



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 179610

TO: Rosanne Kosson  
Location: rem/3B84/3C70  
Art Unit: 1653  
Friday, February 17, 2006  
Case Serial Number: 10/754115

From: Kristine Hensle  
Location: Biotech-Chem Library  
REM-1B69  
Phone: (571) 272-4161

Kristine.Hensle@uspto.gov

### Search Notes

Examiner Kosson,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle  
Librarian  
STIC Biotech/Chem Library  
(571)272-4161

179610  
mej

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 87.0796 Seconds  
(without alignments)  
4606.314 Million cell updates/sec

Title: US-10-754-115-47  
Perfect score: 5005  
Sequence: 1 MKNIDPKLYQKTPVSVYDN.....DAEISFLTPIPLKNVPHKR 960

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5005	100.0	960	US-10-706-424-12	Sequence 12, Appl
2	5005	100.0	960	US-10-754-115-47	Sequence 47, Appl
3	2879	57.5	938	US-10-706-424-16	Sequence 16, Appl
4	2850.5	57.0	949	US-10-706-424-14	Sequence 14, Appl
5	2850.5	57.0	949	US-10-754-115-64	Sequence 64, Appl
6	2740	54.7	1043	US-10-262-794A-61	Sequence 61, Appl
7	2537.5	50.7	915	US-09-817-514A-6	Sequence 6, Appl
8	2535.5	50.7	915	US-10-754-115-58	Sequence 58, Appl
9	2304.5	46.0	962	US-10-754-115-51	Sequence 51, Appl
10	2304.5	46.0	962	US-11-020-848-4	Sequence 4, Appl
11	2226	44.5	1016	US-10-753-901-16	Sequence 16, Appl
12	2226	44.5	1016	US-10-754-115-16	Sequence 16, Appl
13	2184	43.6	973	US-10-754-115-61	Sequence 61, Appl
14	1581.5	31.6	930	US-10-609-113-13	Sequence 13, Appl
15	1581.5	31.6	930	US-10-754-115-42	Sequence 42, Appl
16	1581.5	31.6	953	US-10-609-113-19	Sequence 19, Appl
17	1581.5	31.6	953	US-10-754-115-43	Sequence 43, Appl
18	1268.5	25.3	940	US-10-365-742-64	Sequence 64, Appl
19	651	13.0	286	US-10-609-113-41	Sequence 41, Appl
20	339.5	6.8	2364	US-10-156-761-7834	Sequence 7834, Ap
21	332.5	6.6	932	US-10-282-122A-44617	Sequence 44617, A
22	329.5	6.6	1572	US-10-282-122A-69415	Sequence 69415, A
23	321	6.4	1530	US-10-282-122A-68175	Sequence 68175, A
24	318	6.4	1385	US-10-282-122A-68242	Sequence 68242, A
25	317	6.3	1565	US-10-282-122A-44558	Sequence 44558, A
26	315	6.3	1317	US-10-369-493-11243	Sequence 11243, A
27	307.5	6.1	843	US-10-282-122A-76214	Sequence 76214, A

ALIGNMENTS

RESULT 1

US-10-706-424-12  
; Sequence 12, Application US/10706424  
; Publication No. US20040103455A1  
; GENERAL INFORMATION:  
; APPLICANT: ffrench-Constant, Richard  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens  
; FILE REFERENCE: 62878  
; CURRENT APPLICATION NUMBER: US/10/706,424  
; CURRENT FILING DATE: 2003-11-12  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 12  
; LENGTH: 960  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-10-706-424-12

Query Match 100.0%; Score 5005; DB 4; Length 960;  
Best Local Similarity 100.0%; Pred. No. 1.6e-315;  
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKNIDPKLYQKTPVSVYDN	RGLIIRNIDFHR	TANGDPDTRITRHQYDIHGLNQSIDP	60
Db	1	MKNIDPKLYQKTPVSVYDN	RGLIIRNIDFHR	TANGDPDTRITRHQYDIHGLNQSIDP	60
Qy	61	RLYEAKQTNNTIKPNFLWQYDLTG	NPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR	120	
Db	61	RLYEAKQTNNTIKPNFLWQYDLTG	NPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR	120	
Qy	121	QYETSSILPGRLLSVAEQTPPEEKTSRITERLI	WAGNTEAEKDHNLAGQCVRRHYDTAGVTRL	180	
Db	121	QYETSSILPGRLLSVAEQTPPEEKTSRITERLI	WAGNTEAEKDHNLAGQCVRRHYDTAGVTRL	180	
Qy	181	ESLSLTGTVLSQSSQLLIDTQ	EANNWGNQNETWQNNLADDIYTLTSTFDTATGALLTQTD	240	
Db	181	ESLSLTGTVLSQSSQLLIDTQ	EANNWGNQNETWQNNLADDIYTLTSTFDTATGALLTQTD	240	
Qy	241	KGNIQRLAYDVAGQLNGSWLTKGQTEQVI	IKSLTYSAAQKLRHEHGNVDITEYSYEPE	300	
Db	241	KGNIQRLAYDVAGQLNGSWLTKGQTEQVI	IKSLTYSAAQKLRHEHGNVDITEYSYEPE	300	
Qy	301	TQRLIGIKTRRPSDTKVLQDLREYDPVGNVI	SIRNDAEATRFWNQKMPENTYDLSL	360	
Db	301	TQRLIGIKTRRPSDTKVLQDLREYDPVGNVI	SIRNDAEATRFWNQKMPENTYDLSL	360	
Qy	361	YOLISATGREMANIQSQSHQPPSPAL	PSDNNNTYNTYTRYTYDRCGNLTQIHSHPATQN	420	
Db	361	YOLISATGREMANIQSQSHQPPSPAL	PSDNNNTYNTYTRYTYDRCGNLTQIHSHPATQN	420	

QY 421 NYTNTIVSNRNLVSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480  
 Db 421 NYTNTIVSNRNLVSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480  
 QY 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTRVTVLPLNLELRLTQNSTATTEDL 540  
 Db 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTRVTVLPLNLELRLTQNSTATTEDL 540  
 QY 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600  
 Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600  
 QY 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660  
 Db 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660  
 QY 661 LYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720  
 Db 661 LYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720  
 QY 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSSTTST 780  
 Db 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSSTTST 780  
 QY 781 NLQKSFYLYRADNRSFEEMQSKPEPKANTPLDTKMAROFASIFIGKDTSNLPKETV 840  
 Db 781 NLQKSFYLYRADNRSFEEMQSKPEPKANTPLDTKMAROFASIFIGKDTSNLPKETV 840  
 QY 841 KNIISWGAAPKPKLKDLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 Db 841 KNIISWGAAPKPKLKDLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 QY 901 GQKNLPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLTTIPLKNVKPKHR 960  
 Db 901 GQKNLPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLTTIPLKNVKPKHR 960

RESULT 2

US-10-754-115-47  
 ; Sequence 47, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintrim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Ni, Weiting  
 ; APPLICANT: Zhu, Baolong  
 ; APPLICANT: Merlo, Don  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Meade, Thomas  
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
 ; FILE REFERENCE: DAS-104XC1  
 ; CURRENT APPLICATION NUMBER: US/10/754,115  
 ; PRIOR FILING DATE: 2004-01-07  
 ; PRIOR APPLICATION NUMBER: US 60/441,723  
 ; PRIOR FILING DATE: 2003-01-21  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 47  
 ; LENGTH: 960  
 ; TYPE: PRT  
 ; ORGANISM: Photorhabdus luminescens  
 US-10-754-115-47

Query Match 100.0%; Score 5005; DB 4; Length 960;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-315;  
 Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHRITANGDPDTRITRHOYDIGHLHQSIDP 60  
 QY 61 RLYEAKQTNNTIKNFLLWQYDLTGNPCTESIDAGRTVTLNDIEGRPLLTATGCVIOTR 120  
 Db 61 RLYEAKQTNNTIKNFLLWQYDLTGNPCTESIDAGRTVTLNDIEGRPLLTATGCVIOTR 120  
 QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLLWAGNTAEKDHNLGAGQCVHYDAGVTRL 180  
 Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLLWAGNTAEKDHNLGAGQCVHYDAGVTRL 180  
 QY 181 ELSLGTGTVLSQSSQLLIDTQEANWTDNETVWQNLADDIYTTLTSTFDATGALLTQDA 240  
 Db 181 ELSLGTGTVLSQSSQLLIDTQEANWTDNETVWQNLADDIYTTLTSTFDATGALLTQDA 240  
 QY 241 KGNITQRLAYDVAGQINGSWLTKGQTEQVVIKSLTYSAGQKLREHNDVITEYSPE 300  
 Db 241 KGNITQRLAYDVAGQINGSWLTKGQTEQVVIKSLTYSAGQKLREHNDVITEYSPE 300  
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAAEATRFWLNQKVPENTTYDLS 360  
 Db 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAAEATRFWLNQKVPENTTYDLS 360  
 QY 361 YQLISATGREMANICQOSHQPSPALPSDNNNTYNTYTRTYTYDRGNLTKIOHSSPATON 420  
 Db 361 YQLISATGREMANICQOSHQPSPALPSDNNNTYNTYTRTYTYDRGNLTKIOHSSPATON 420  
 QY 421 NYTNTIVSNRNLVSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480  
 Db 421 NYTNTIVSNRNLVSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480  
 QY 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTRVTVLPLNLELRLTQNSTATTEDL 540  
 Db 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTRVTVLPLNLELRLTQNSTATTEDL 540  
 QY 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600  
 Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600  
 QY 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660  
 Db 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660  
 QY 661 LYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720  
 Db 661 LYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720  
 QY 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSSTTST 780  
 Db 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSSTTST 780  
 QY 781 NLQKSFYLYRADNRSFEEMQSKPEPKANTPLDTKMAROFASIFIGKDTSNLPKETV 840  
 Db 781 NLQKSFYLYRADNRSFEEMQSKPEPKANTPLDTKMAROFASIFIGKDTSNLPKETV 840  
 QY 841 KNIISWGAAPKPKLKDLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 Db 841 KNIISWGAAPKPKLKDLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 QY 901 GQKNLPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLTTIPLKNVKPKHR 960  
 Db 901 GQKNLPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLTTIPLKNVKPKHR 960

RESULT 3

US-10-706-424-16  
 ; Sequence 16, Application US/10706424  
 ; Publication No. US20040103455A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ffrench-Constant, Richard  
 ; APPLICANT: Waterfield, Nicholas  
 ; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens  
 ; FILE REFERENCE: 62878



Qy	1	MKNIDPKLYQKTPVTSVYDNRGLIIRNIDFHRTTANGDDPTRIIRTHQVYDIHGLNQSIDP	60
Db	1	MENIDPKLYHHTPTVSVHDNRGLAIRNISFRTTAEANTDTRIIRTHQVYDNRGLNQSIDP	60
Qy	61	RLYEAKOTNNITIKENFLWOYDLTCNPLCTESIDAGRTVTLNDISGRPLLTWTATGVICTR	120
Db	61	RLYDAKOTNNAVQPNFIRWHMLTGNILKTESVDAGRTITLNDIISGRPVLTITNAAGVRQNH	120
Qy	121	QYETSSLPGRLLSVAEQ--TPBEKTSRIITERLIWAGNTEAEKXHNLAGOCVRHYDTAGVTR	179
Db	121	RYEDNTLPGRLLAISQEQQAEBKT--TERLIWAGNTPQEKXHNLAGOCVRHYDTAGLTQ	177
Qy	180	LESLSLTCVTLSQSSOLLITQEBANWTGDNETVWQNLADDIYTLTFTPDATGALLTQTD	239
Db	178	LNSLALTGAVLSQSOQLLTDNQDADWTGDSQWLQOQLSSDVIYITQSNMTDATGALLTQTD	23
Qy	240	AKGNIQRLAYDVACQLANGSMWLTAKGQTEQVILKSLITYSAAQCKLREHGNDVITEYSYEP	299
Db	238	AKGNIQRLAYDVACQLANGSMWLTAKGAQEQVILKSLITYSAAQCKLREHNGIVITEYSYEP	297
Qy	300	ETQRLIGIKTRRPSDTKVQLDRYEDVPVGNVISIRINDAEATRFWKNQKWPENTYITYDS	359
Db	298	ETQRLIGITTRRPSDAKVQLDRYQYDPVGNVISIRINDAEATRFWRNQKVAPENSITYDS	357
Qy	360	LYQLISATGREMANIGQOSHOPSPALPSPDNNTYNTYTRITYYDRGNGNLTKIOHSSPATQ	419
Db	358	LYQLISATGREMANIGQSNQLPSPALPSPDNNTYNTYTRITYYDRGNGNLTKIOHSSPAQ	417
Qy	420	NNYTNTITSNRSNRVLSLTLEDPQAQVDALFDAGGHQNTLISGONLWNTRGBELQOQVTL	479
Db	418	NNYTNTITSNRSNRVLSLTLTADPTQVDALFDAGGHQTSLLSGQVLTWTPRGLBKQ---	474
Qy	480	VKRDKGANDD--REWYRSGDGRWMLKINEOQASNNACTORAVTYLPNLELRTNSTATT	537
Db	475	-----ANNAGNENWYRYSNGIQLKVNQEQOTQNIPOQQRVITYLPGLIEHRTQNNATT	528
Qy	538	EDLQVITVGBAGRAQVRVLHWESCKPEDIDNNQLRYSYDNLIGSSOLELDSGQIISBEE	597
Db	529	BELHVIITLKGAGRAQVRVLHWESCKPEDINNQLRYSYDNLIGSSQLQDSDGQIISBEE	588
Qy	598	YYPYGGTALWAARNQOTASXYTIIRYSGKERDATGLYYGYRYYQPWIRGLWSSDPAGTID	657
Db	589	YYPYGGTALWAARNQOTASXYTIIRYSGKERDVTLYYGYRYYQPWAGRWLGADPAGTID	648
Qy	658	GLNLYMRVNRNPTVLLDPDGLMPTIAERIAALKKNKVTDSPANATVAI-----N	710
Db	649	GLNLYMRVNRNPTVQFDVQGL-----SPANRTEAIIKQGSFTG	687
Qy	711	IRPPVAPKPSLPK-----ASTSQP-----TTHPIGAIKIP-----TTSGSSIVAPL	753
Db	688	MEBAVYTKMAKPTQFKQRAITAAQTEOEAHESLTNNP--SVDISPIKNYITTTDSSQINAAI	745
Qy	754	SPVGNKSTSEISLPSAQSSSSSTSTNLQKSFTRYADNRSPFEQSKPEPFKAWTP	813
Db	746	RE--NRITPAV---ESLDATLSSLDQDMRVYRWMTVDNS-----TP	784
Qy	814	LDTKMARQFASIFIGQKDTSNLPKETVKNTISWGAKPKLDLSNYIKYTKDKSTVWVSTA	873
Db	785	SPWHISPOEGHSINVGDIIVSDNAYLSTSAH-----RGLFNFKCKETSETRYVQMA	834

```
QY 599 YPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 YPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
QY 659 LNLRYMVRNNPVTLLDPDGLMPTTAERIAALKKN-----KVTDSAPSPANATNVAIN 710
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 LNLRYMVRNNPVTLLDPDGLMPTTAERIAALKKN-----KVTDSAPSPANATNVAIN 710
QY 711 IR-----PPV-----APKPSL---PKASTSSQPTTHPIGAANIK---PTTSGSIVAPLSPV 756
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 711 DEFLHYGVPDKLBEKKPGLNVPBELFDRGPSENGVSTLTFFKDLPIPCISINTBYTLDIL 770
QY 757 GNKSTSEISLSPESAQSSSSSTTSLNLOKSKFTLYRADNRSPFEEMOSK-FPBGFKAWTPLD 815
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 771 YNKHETK-PFYPENEATVGADLGV-IMSVFPGNKSIGNADEDLKEHLPLGKSTMDKTD 828
QY 816 TKMARQFASIFIGOKDTSNL-----PKETVKNISTWGAKPKLKLDSNY 858
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 829 LPDLKQ--GLMIAEKIKSGKGNYPFHGAAIAVVVGEDKKVAASILTDLSEPK-RDEGEY 885
QY 859 IKYTKOKSTVWVSTAINTEAGQSSGAPLHKIDMDLYEFA-IDGQKLNPLPEGRTKN 914
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 886 LQSTRKVSAMEI-TNVNEFRG-----HDYPKSKYSIGLVTAERKQPVISKKRAN 933

RESULT 5
US-10-754-115-64
; Sequence 64, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 64
; LENGTH: 949
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-64

Query Match 57.0%; Score 2850.5; DB 4; Length 949;
Best Local Similarity 60.9%; Pred. No. 8.1e-176;
Matches 583; Conservative 100; Mismatches 207; Indels 67; Gaps 17;

QY 1 MKNIDPKLYQKTPVSVVDNRGLIRNIDFRTTANGDPDTRITRHOYDIGHLNSQSIDP 60
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MKNIDPKLYQHTPTVNVVDNRGLIRNIDFHRVAGGDTTRITRHOYDTRGHLSQSIDP 60
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 61 RLYEAKQNTNTIKPNFLWOYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTARGVIQTR 120
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RLYDAKQNTNSTPNFLQYNLTGDLTRESVDAGRTVALNDIEGRQLIVTATGAIQTR 120
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 QYESSLPGRLLSVAEQTPBEKTSITELLIWAGNTEAKDNLGACVRYHVDTAGVTRL 180
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 QYEAENTLPGRLLSVSEQAPGQTPKVTSEHFWAGNTQAEKDNLAGQVRYHVDTAGVTRL 180
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 ESSLTGTGVLSSQSLLDITQEAQNTWGNQNTWQNLADDIYTLTSTFDATGALLTQDTA 240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Db 181 ESSLTENTILSSQSQLLADGQADWTGNDETLWQTKLNSETYTTQSTFDATGALLTQDTA 240
QY 241 KGNLQRLAYVAGQINGSWLTKGQTEQVILKSLTYSAAQOKLREHGNVDVITSYSEPE 300
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 KGNMQLAYNAVAGLOQSGSWLTKNQSQVIVKSLTYSAAQOKLREHGNGVITSYSEPE 300
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 TORLIGIKTRRPSDKVLQDLRYEYDPVGNVISIRNDAEATRFWNHOKWMPENTYTYDLSL 360
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 TRLRIGITTRRQSDSKVLQDLRYEYDPVGNVISIRNDAEATRFWNHOKWMPENTYTYDLSL 360
QY 361 YQLISATGREMANIQOOSHQPSP--ALPSDNNNTYNTYTYTYDRGNLTKIOHSPAT 418
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 YQLISATGREMANIQOOSNQLPSPIIPLPTDENSYNTYTRSYNDYDRGNLVOIRHSPAA 420
QY 419 QNNYTTITVSNRNRVAVLSTLTPAQVDALPAGGHONTLISGQNLNNTRGELQOVT 478
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 QNNYTTITVSNRNRVAVLSSLTSDPTQVEALPAGGHQTKLLPGQELSMNTRGELQOVT 480
QY 479 LVKRDKGANDREWYRSGDGRMLKINEQOASNAQOTQRTVYLPNLELRLTQNSTATTE 538
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 PVSEAS--DREWYRYGNDGMRELKYSQOQTGNSQOQRTVYLPDLRLTTQNGTTISE 538
QY 539 DLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSQIISBEY 598
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 DLHAIITVGAAGHAQVRVLHWETTPPAGINNQLRYSYDNLIGSSQLELDNAGQIISBEY 598
QY 599 YPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 YPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
QY 659 LNLRYMVRNNPVTLLDPDGLMPTTAERIAALKKN-----KVTDSAPSPANATNVAIN 710
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 LNLRYMVRNNPVTLLDPDGLMPTTAERIAALKKN-----KVTDSAPSPANATNVAIN 710
QY 711 IR-----PPV-----APKPSL---PKASTSSQPTTHPIGAANIK---PTTSGSIVAPLSPV 756
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 711 DEFLHYGVPDKLBEKKPGLNVPBELFDRGPSENGVSTLTFFKDLPIPCISINTBYTLDIL 770
QY 757 GNKSTSEISLSPESAQSSSSSTTSLNLOKSKFTLYRADNRSPFEEMOSK-FPBGFKAWTPLD 815
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 771 YNKHETK-PFYPENEATVGADLGV-IMSVFPGNKSIGNADEDLKEHLPLGKSTMDKTD 828
QY 816 TKMARQFASIFIGOKDTSNL-----PKETVKNISTWGAKPKLKLDSNY 858
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 829 LPDLKQ--GLMIAEKIKSGKGNYPFHGAAIAVVVGEDKKVAASILTDLSEPK-RDEGEY 885
QY 859 IKYTKOKSTVWVSTAINTEAGQSSGAPLHKIDMDLYEFA-IDGQKLNPLPEGRTKN 914
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 886 LQSTRKVSAMEI-TNVNEFRG-----HDYPKSKYSIGLVTAERKQPVISKKRAN 933

RESULT 6
US-10-262-794A-61
; Sequence 61, Application US/10262794A
; Publication No. US20030207806A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petelli, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhupinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
```

Query Match	54.7%	Score 2740;	DB 4;	Length 1043;
Best Local Similarity	74.7%;	Pred. No. 1.4e-168;		
Matches 513;	Conservative	78;	Mismatches 88;	Indels 8; Gaps 3;
QY	1	MKKIDPKLYOKTPTVSVDNRGLIIRNIDFRTTANGDDPTRTTRHQYDIHGLNOSIDP	60	
DB		1	MSPSYTLTYQTPTVSVDNRGLSIRIDGFRIVIGGDDTTRVTRHQYDARGHLNYSIDP	60
QY	61	RLYEAKQTNNTIKPFLWQYDLDTGNPLCTESIDAGRVTTLNDIEGRPLLTVTATGVQTR	120	
DB	61	RLYDAKQADNSVKNPFVQKQDLGHALRTESVDAGRVTVALNDIEGRSVMTWMTATGVQTR	120	
QY	121	QYETSSLPGRLLSVAGOTPEKTSRITRERLIWAGNTEAEKHNLAGOCVRHYDTAGVTRL	180	
DB	121	RYEGNTLPGRLLSVSEQVFNQESAKVTERFIWAGNTTSEKYNLSGCI RHYDTAGVTRL	180	
QY	181	ESLSLCTGLVSQSOLLIDTQEANWTGDNETWQNMALDITTLTSLTFDATTGALLTQDA	240	
DB	181	MSQSLAGAMLSQSHLLAEGQEANWSGDDTETVQCMCLASEVYTTQSTTNAIGALLTQDA	240	
QY	241	KGNIRIAYDVAGOLNGSWLTLKGQTEQVFIKSLTYSAAQOKLRBEHGNVDITEYSYEPE	300	

Query Match	50.7%	Score 2537.5	DB 3	Length 915
Best Local Similarity	70.1%	Pred. No. 1.5e-155		
Matches 488	Conservative 75	Mismatches 108	Indels 25	Gaps 6
Qy	1	MKNIDPKLYOKTPTVSVDNRGLIIRNIDFHRITTTANGDPDTRIFRHOVDIHGLNQSIDP	60	
Db	1	MSYNSAIDOKTSPSIKVLNDRKLNVRLEVLRTQADENSDELITFYEFNIPGVQVKSTDP	60	
Qy	61	RLYEAKOTNNITIKPNFLQVYDLTGPNLCITSEIDAGRTVTNLNDIEGRPLLTTVATGVIQTR	120	
Db	61	R-----KKNQSGPNFRIVENLAGQVLRSEESVDAGRTITLNDIEBSPVLIIINATGVRQNH	115	
Qy	121	QYETSSILPGRLLSVAECTPE-EKTSRITRTERLIWAGNTEAEKDHNLACQVVRHYDTAGVTR	179	
Db	116	RYEDNITLPGKLLAITQVQAGEKTI---TERLIWAGNTPQEKDYNLACQVVRHYDTAGTQ	172	
Qy	180	LESLSLTGTVLSSQSOLLIDTQBANMTGDNETVWQNNLADDIYVTLSTFDATGALLTQTD	239	
Db	173	LNLSILAGVVISQSOLLITNDQADMTGDSQWLWQKLSDDVYITQSNLTATGALLTQTD	232	

Query Match	50.7%	Score 2537.5	DB 3	Length 915
Best Local Similarity	70.1%	Pred. No. 1.5e-155		
Matches 488	Conservative 75	Mismatches 108	Indels 25	Gaps 6
Qy	1	MKNIDPKLYOKTPTVSVDNRGLIIRNIDFHRITTTANGDPDTRIFRHOVDIHGLNQSIDP	60	
Db	1	MSYNSAIDOKTSPSIKVLNDRKLNVRLEVLRTQADENSDELITFYEFNIPGVQVKSTDP	60	
Qy	61	RLYEAKOTNNITIKPNFLQVYDLTGPNLCITSEIDAGRTVTNLNDIEGRPLLTTVATGVIQTR	120	
Db	61	R-----KKNQSGPNFRIVENLAGQVLRSEESVDAGRTITLNDIEBSPVLIIINATGVRQNH	115	
Qy	121	QYETSSILPGRLLSVAECTPE-EKTSRITRTERLIWAGNTEAEKDHNLACQVVRHYDTAGVTR	179	
Db	116	RYEDNITLPGKLLAITQVQAGEKTI---TERLIWAGNTPQEKDYNLACQVVRHYDTAGTQ	172	
Qy	180	LESLSLTGTVLSSQSOLLIDTQBANMTGDNETVWQNNLADDIYVTLSTFDATGALLTQTD	239	
Db	173	LNLSILAGVVISQSOLLITNDQADMTGDSQWLWQKLSDDVYITQSNLTATGALLTQTD	232	

```
Qy 240 AKGNIQRLAYDVAGQVLSWTLKQTEQVLIKSLTYSAAQKLEBEHNDVITEYSYEP 299
Db 233 AKGNIQRLAYDVAGQVLSWTLKQAEQVLIKSLTYSAAQKLEBEHNGVITEYSYEP 292
Qy 300 ETQRLIGIKTRRPSDVKVLODLRYEYDPVGNVIRINDAEATRFWRKQKVPENTTYDYS 359
Db 293 ETQRLIGIKTRRPSDAKVLQDLRYQYDPVGNVINIRINDAEATRFWRKQKVPENSYYDYS 352
Qy 360 LYQLISATGREMANIGQSHQFSPALPDSNNNTYTRTYDRGGLTKIHHSSPATQ 419
Db 353 LYQLISATGREMANIGQNNQLPSPALPDSNNNTYTRSYDHSNGLTIQRHSSPATQ 412
Qy 420 NNYTNTITVNSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQNLNWNTRGELQVTL 479
Db 413 NNYTVAITLNSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQNLNWNTRGELQVTL 470
Qy 480 VKRDKGANDREWRYSGDGRMLKINEQQAASNAQTORVTLNLELRLTONSTATTED 539
Db 471 ---NNGPGN--EWTRYDSNGMRQLKVEQPTQNTQQORVYIYLPGLRLTQSNATTEE 525
Qy 540 LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEQIISEBEYY 599
Db 526 LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELDNQOQIISEBEYY 585
Qy 600 PYGTALWAARNQTEASYKTIYRSGKERDATGLYYYGYRYYPQWIGRWLSDPAGTIDGL 659
Db 586 PFGGTALWAANSQTEASYKTIYRSGKERDATGLYYYGYRYYPQWAGRWLSADPAGTIDGL 645
Qy 660 NLYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVT 695
Db 646 NLYRMVRNPNVSLQDENG LAP-----EKGYT 672
```

## RESULT 8

```
US-10-754-115-58
; Sequence 58, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR FILING DATE: 2003-01-21
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 58
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-58
```

```
Query Match 50.7%; Score 2535.5; DB 4; Length 915;
Best Local Similarity 70.1%; Pred. No. 2e-155;
Matches 488; Conservative 76; Mismatches 107; Indels 25; Gaps 6;

Qy 1 MKNIDPKLYQKTPVSVVDNREGLIRNIDFRTTANGDPDTRITRHHQYDIGHLNQSIDP 60
Db 1 MSSYNSAIDQKTPSIKVLNDNKNJNVRLTLEYLRQADNSDELITFYEFNIFGVQKSTDP 60
Qy 61 RLYEAKQTNNTKFNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
```

```
Db 61 R-----KKNQSGPNFRVFNLAGQVLRSESDAGRTITLNDIESRPVLINATGVQRNH 115
Qy 121 QYETSSLPGBLLSVABQTPB- EKTSRITERLIWAGNTEAEKDHNLACQCVRHYYTAGVTR 179
Db 116 RYEDNTLPGRLLATTEQVQAGEKT---TERLIWAGNTPQEKDYNLAGQCVRHYYTAGLTQ 172
Qy 180 LESLSLTGTVLSOSSQLLIQTQANWTGDNETVWQNNLADDIYTLTSTFEDATGALLTQTD 239
Db 173 LNSLSLAVGVLSOSSQLLVDDKNADWTGDSQKLSQSDVYITQKADATGALLTQTD 232
Qy 240 AKGNIQRLAYDVAGQVLSWTLKQTEQVLIKSLTYSAAQKLEBEHNDVITEYSYEP 299
Db 233 AKGNIQRLAYDVAGQVLSWTLKQAEQVLIKSLTYSAAQKLEBEHNGVITEYSYEP 292
Qy 300 ETQRLIGIKTRRPSDVKVLODLRYEYDPVGNVIRINDAEATRFWRKQKVPENTTYDYS 359
Db 293 ETQRLIGIKTRRPSDAKVLQDLRYQYDPVGNVINIRINDAEATRFWRKQKVPENSYYDYS 352
Qy 360 LYQLISATGREMANIGQSHQFSPALPDSNNNTYTRTYDRGGLTKIHHSSPATQ 419
Db 353 LYQLISATGREMANIGQNNQLPSPALPDSNNNTYTRSYDHSNGLTIQRHSSPATQ 412
Qy 420 NNYTNTITVNSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQNLNWNTRGELQVTL 479
Db 413 NNYTVAITLNSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQNLNWNTRGELQVTL 470
Qy 480 VKRDKGANDREWRYSGDGRMLKINEQQAASNAQTORVTLNLELRLTONSTATTED 539
Db 471 ---NNGPGN--EWTRYDSNGMRQLKVEQPTQNTQQORVYIYLPGLRLTQSNATTEE 525
Qy 540 LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEQIISEBEYY 599
Db 526 LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELDNQOQIISEBEYY 585
Qy 600 PYGTALWAARNQTEASYKTIYRSGKERDATGLYYYGYRYYPQWIGRWLSDPAGTIDGL 659
Db 586 PFGGTALWAANSQTEASYKTIYRSGKERDATGLYYYGYRYYPQWAGRWLSADPAGTIDGL 645
Qy 660 NLYRMVRNPNVTLDDPGLMPTIAERIAALKKNKVT 695
Db 646 NLYRMVRNPNVSLQDENG LAP-----EKGYT 672
```

## RESULT 9

```
US-10-754-115-51
; Sequence 51, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR FILING DATE: 2003-01-21
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 51
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-10-754-115-51
```

Query Match 46.0%; Score 2304.5; DB 4; Length 962;  
Best Local Similarity 64.6%; Pred. No. 2.1e-140;  
Matches 445; Conservative 94; Mismatches 145; Indels 5; Gaps 4;

QY 1 MKNIDPKLYOKTPTVSVVVDNRGLIIRNIDFHTTANGDPDTRITRPHQYDIHGLNQSDP 60  
DB 1 MNVFNPTLYAGTPTVTVMNDRGLSVRDIAVHRTTAGEQADTRITRHOQYSPHNFLESIDP 60

QY 61 RLYEAKQTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIOTR 120  
DB 61 RLFDL-QSQSTIKPNFTYCPALKGDLVATESVDAGQTVILSDIEGRPLLNISAMGVVXHW 119

QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRYHYDTAGVTRL 180  
DB 120 QYESTLPGRLLAVERKNEASTPQIIERFIWSGNSPSEKDHNLGAGLYRHYDTAGLNOL 179

QY 181 ELSLSITGVLSQSSQLLIDTQENWNTGNETVQNMALADDIYTLTSTFDATGALLTQDA 240  
DB 180 NAVSLTSVDLSQSRLQLQDDVTADWSGDSQWKTRLSNDIFTTBITADAVGNFLTQDA 239

QY 241 KGNTRQLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHNDVITEYSPE 300  
DB 240 KSNQRLSYDVAGQLKASWLTIKGNEQVIVNSLTYSAGOKLREBQGVVTEYSYEAQ 299

QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVGNVISIRNDAEATRFWNNQKMPENTTYDLSL 360  
DB 300 TWRLIGVTAYRQSKKRLQDLVYNDPVGNLIRNNAEATRFWRNQIIVEPENHYAYDSL 359

QY 361 YOLISATGREMANTQOQSHQPPSP--ALPSDNNTVNTYTRTYHYDRGNLTKIQHSSPAT 418  
DB 360 YOLISASGREIASIGQQGSRPLPPIPLPANDDVVYTRTYHYDRGNLTKQIRHCAPAT 419

QY 419 QNNYTNITVNSNRNAVLSLTEDPAQVDALFDAGGHQNTLISQNLNMTNRGELQVIT 478  
DB 420 DNKYTKITVNSNRNAVWLTITDPAKVDITLFDHGHQLQSQGTLQCNWTRGELQIT 479

QY 479 LVKRDKGANDREWRYSGDGRMLKINEQASNNAAQTVTYLPNLELRTQNSTATTE 538  
DB 480 KIQRDE-KPADKERYRYGVGAARVVKISTQOAGSSSHVQVRVYLPGLRLTTOHDA TLIE 538

QY 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSOLELDSGQIIISBEY 598  
DB 539 DLQVIIMGEAGRAQVRVLHWEIPPPDNLNLSRLSYSDLSMGSSQLELDGAGQIITQBEY 598

QY 599 YPYGCTALWAARNQTEASYKTIYRSGKERDATGLYYGYRYQYQPMIGRWLSADPAGTIDG 658  
DB 659 LNLRYMVRNPNITVRDADGLAP-IGDKIS 686

RESULT 10  
US-11-020-848-4  
; Sequence 4, Application US/11020848  
; Publication No. US20050155104A1  
; GENERAL INFORMATION:  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Hey, Timothy  
; APPLICANT: Sheets, Joel  
; APPLICANT: Meade, Tom  
; APPLICANT: Li, Ze Sheng  
; APPLICANT: Lira, Justin  
; APPLICANT: Russell, Sean  
; APPLICANT: Thompson, Robin  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Fencil, Kristin  
; TITLE OF INVENTION: Toxin Complex Proteins and Genes from Xenorhabdus bovienii  
; FILE REFERENCE: DAS-114XCI  
; CURRENT APPLICATION NUMBER: US/11/020,848  
; CURRENT FILING DATE: 2004-12-23

; PRIOR APPLICATION NUMBER: US 60/534,893  
; PRIOR FILING DATE: 2004-01-07  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 962  
; TYPE: PRT  
; ORGANISM: Xenorhabdus bovienii  
US-11-020-848-4

Query Match 46.0%; Score 2304.5; DB 6; Length 962;  
Best Local Similarity 64.6%; Pred. No. 2.1e-140;  
Matches 445; Conservative 94; Mismatches 145; Indels 5; Gaps 4;

QY 1 MKNIDPKLYOKTPTVSVVVDNRGLIIRNIDFHTTANGDPDTRITRPHQYDIHGLNQSDP 60  
DB 1 MNVFNPTLYAGTPTVTVMNDRGLSVRDIAVHRTTAGEQADTRITRHOQYSPHNFLESIDP 60

QY 61 RLYEAKQTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIOTR 120  
DB 61 RLFDL-QSQSTIKPNFTYCPALKGDLVATESVDAGQTVILSDIEGRPLLNISAMGVVXHW 119

QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRYHYDTAGVTRL 180  
DB 120 QYESTLPGRLLAVERKNEASTPQIIERFIWSGNSPSEKDHNLGAGLYRHYDTAGLNOL 179

QY 181 ELSLSITGVLSQSSQLLIDTQENWNTGNETVQNMALADDIYTLTSTFDATGALLTQDA 240  
DB 180 NAVSLTSVDLSQSRLQLQDDVTADWSGDSQWKTRLSNDIFTTBITADAVGNFLTQDA 239

QY 241 KGNTRQLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHNDVITEYSPE 300  
DB 240 KSNQRLSYDVAGQLKASWLTIKGNEQVIVNSLTYSAGOKLREBQGVVTEYSYEAQ 299

QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVGNVISIRNDAEATRFWNNQKMPENTTYDLSL 360  
DB 300 TWRLIGVTAYRQSKKRLQDLVYNDPVGNLIRNNAEATRFWRNQIIVEPENHYAYDSL 359

QY 361 YOLISATGREMANTQOQSHQPPSP--ALPSDNNTVNTYTRTYHYDRGNLTKIQHSSPAT 418  
DB 360 YOLISASGREIASIGQQGSRPLPPIPLPANDDVVYTRTYHYDRGNLTKQIRHCAPAT 419

QY 419 QNNYTNITVNSNRNAVLSLTEDPAQVDALFDAGGHQNTLISQNLNMTNRGELQVIT 478  
DB 420 DNKYTKITVNSNRNAVWLTITDPAKVDITLFDHGHQLQSQGTLQCNWTRGELQIT 479

QY 479 LVKRDKGANDREWRYSGDGRMLKINEQASNNAAQTVTYLPNLELRTQNSTATTE 538  
DB 480 KIQRDE-KPADKERYRYGVGAARVVKISTQOAGSSSHVQVRVYLPGLRLTTOHDA TLIE 538

QY 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSOLELDSGQIIISBEY 598  
DB 539 DLQVIIMGEAGRAQVRVLHWEIPPPDNLNLSRLSYSDLSMGSSQLELDGAGQIITQBEY 598

QY 599 YPYGCTALWAARNQTEASYKTIYRSGKERDATGLYYGYRYQYQPMIGRWLSADPAGTIDG 658  
DB 599 YPYGCTALWAARNQTEANYKTIYRSGKERDATGLYYGHYRYPQPMIGRWLSADPAGTIDG 658

QY 659 LNLRYMVRNPNITVRDADGLAP-IGDKIS 687  
DB 659 LNLRYMVRNPNITVRDADGLAP-IGDKIS 686

RESULT 11  
US-10-753-901-16  
; Sequence 16, Application US/10753901  
; Publication No. US20040194164A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Larrinua, Ignacio  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Schafer, Barry



Db 590 SDGQIIISQEEYYPYGGTAVMAARSQSSEADYKTVRYSGKERDATGLYYYGYRYYSQSWTGRW 649  
Qy 648 LSSDPAGTIDGLNLYRMVRNPNVTLLDPDGLMP 680  
Db 650 LSVDPAGEVDGLNLFMRNPNPIVPSDSGGRPP 682

RESULT 13  
US-10-754-115-61  
; Sequence 61, Application US/10754115  
; Publication No. US20040208907A1  
; GENERAL INFORMATION:  
; APPLICANT: Hey, Timothy  
; APPLICANT: Schleper, Amanda  
; APPLICANT: Bevan, Scott  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Li, Ze Sheng  
; APPLICANT: Ni, Weiting  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Don  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Meade, Thomas  
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
; FILE REFERENCE: DAS-104XC1  
; CURRENT APPLICATION NUMBER: US/10/754,115  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR APPLICATION NUMBER: US 60/441,723  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 61  
; LENGTH: 973  
; TYPE: PRT  
; ORGANISM: Serratia entomophila  
US-10-754-115-61

Query Match 43.6%; Score 2184; DB 4; Length 973;  
Best Local Similarity 60.5%; Pred. No. 1.4e-132;  
Matches 435; Conservative 87; Mismatches 149; Indels 48; Gaps 9;

Qy 8 LYQKTPVSVVDNRGLIIRNIDFHTTANGDPDT-----RITHQYDIHGLNQSIDPR 61  
Db 5 LFSSTPSVAVLDNRGLLRQLYR-----HPDTPETDERITCQHDERSLSQSADPR 59  
Qy 62 LYBAKQTNNTIKPNFLWQYDLTGPNLCSTESIDAGRTVTLNDIEGRPLLTATG---VI 117  
Db 60 LHAAGLT-----NFTYLSNLTGTVLQSVSDAGTSLSDAAGRAFLAVTGTAGTEDAVT 113  
Qy 118 QTRQYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGQCVRHVDTAGV 177  
Db 114 RTWQYEDDTLPGRLPSITEQVTGE-AAQITERFVYAGNTDAEKILNLAGQCVSHVDTAGL 172  
Qy 178 TRLESLSLTGTVLSSQSLLDTOEANTWGNETWQNMADDIYTTLTSTFDATGALLTQ 237  
Db 173 VQTDLSIALSGVPLAVTROLLPDAAGANWNGDASAWNDLLDGETFFPTQTHADATGAVLSI 232  
Qy 238 TDAGNIORLAYDVAGQLNGSWLTLKGOTEQVVIKSLTYSAGQKLRREHGNVDVTEYSY 297  
Db 233 TDAGNLRQVAYDVAGLLSGSWLTLKGOTEQVVIASLYSAGKLRREHGNVGVTSIY 292  
Qy 298 EPETQRLIGIKTRRPS-----DTKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKWPEN 353  
Db 293 EPETQRLTGIKTERPSGHVAGAKVLQDLRYTDPVGNVLSYNNDAEATRFWRNOKVPEEN 352  
Qy 354 TYTDSLVLQISATGREMANTGQSHQPPSPA--LPSDNNTYNTYTRYTDRGNLTKI 411  
Db 353 TYIYDSLVLQVSATGREMANAGQGNLPSATAPLPTDSSAYNTYTRYTDRGNLTKI 412  
Qy 412 QHSSPATONNTYNTITVSNRNLAVLSTLTEDPAQVDALPDAGGHONTLISGONLWNT 471  
Db 413 RHPAPATNNNTYTDITVSDRSNRNLAVLSTLAEPFSDVDMFLFSGGQKHLQPGQALVWTPR 472

Qy 472 GELQOVLTKRDKGANDREWRYSGDGRRLMKINEQOASNAQTQRTVYILPNLELRITQ 531  
Db 473 GELQKVTVPVRDGA--DDESRYRDAGSORIKTGTQRTGNNVQTRVYVLPGLRLTMA 531  
Qy 532 NSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRVSVDNLLSGSOLELDSGQ 591  
Db 532 NGVTEKSLQVITVGEAGRAQVRVLHWEIGKPDLDDESDVSRYSDNVLGSSQLELDREGY 591  
Qy 592 IISBEEYYPYGGTALWAARNTEASYKTIRYSGKERDATGLYYYGYRYYPWIGRWLSSD 651  
Db 592 LISBEEYYPYGGTAVLTARSEVADYKTIRYSGKERDATGLDYGYRYYPWAGRWLSTD 651  
Qy 652 PAGTIDGLNLYRMVRNPNVTLLDPDGLMPT-----IAERIAALKK 691  
Db 652 PAGTVDGLNLFMRVNRNPNVTLLFDSNGRISTQGEARRLVGEAFVHPLHMPVERISVERK 710

## RESULT 14

US-10-609-113-13  
; Sequence 13, Application US/10609113  
; Publication No. US20040110184A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Bevan, Scott  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Donald J.  
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from  
; TITLE OF INVENTION: Paenibacillus Species  
; FILE REFERENCE: DAS-101XC2  
; CURRENT APPLICATION NUMBER: US/10/609,113  
; CURRENT FILING DATE: 2003-06-27  
; PRIOR APPLICATION NUMBER: US 60/392,633  
; PRIOR FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/441,647  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 13  
; LENGTH: 930  
; TYPE: PRT  
; ORGANISM: Paenibacillus strain IDAS 1529  
US-10-609-113-13

Query Match 31.6%; Score 1581.5; DB 4; Length 930;  
Best Local Similarity 42.2%; Pred. No. 1.5e-93;  
Matches 370; Conservative 128; Mismatches 284; Indels 95; Gaps 23;

Qy 8 LYQKTPVSVVDNRGLIIRNIDFHTTANGDPDTRITEHQYDIHGLNQSIDPRLYEAKQ 67  
Db 6 IYRGTPTISVVVDNRNLEIRILOYNRIAAEDPADECILRNTYTPLSYLGSSMDPRLSQVQ 65  
Qy 68 TNNTIKPNFLWQYDLTGPNLCSTESIDAGRTVTLNDIEGRPLLTATGVIQTRQYETSLL 127  
Db 66 DDRCPTPNIRWASLRGALCSVDAGRAKAEFLDIEGRPVWLDANGTETTLEYD---V 122  
Qy 128 PGRLLSVAEQTPPEKTSRITERLIWAGNTEAE-KDHNLAGQCVRHVDTAGVTRLESLSLT 186  
Db 123 LGRPTAVFEQOEGTDSPOCRRERFIY-GEKEADAQANRLRGQLVRHYDTAGRIQTDLSILA 181  
Qy 187 GTVLSSQSLLDITQE-ANWTGDMETWQNMADDIYTTLTSTFDATGALLTQTDAGNIQ 245  
Db 182 GLPLRQSRQLLKNWDEPDGMSDBESAWASLLAAEAYDTSWRYDAQDRVLAQTDAGNIQ 241  
Qy 246 RLAYDVAGQLNGSWLTLKGOTEQVVIKSLTYSAGQKLRREHGNVDVITEYSYEPETORLI 305  
Db 242 QLTNDAGQPOAVSLKQCGQAEQRIWNRIENAGQVDLAEGAGIVITEYVEESTORLI 301  
Qy 306 GIKTRR---PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKWPENYTYDLSLYQ 362  
Db 302 RKDSRGLSSGEREVLDQRYEYDPVGNILSIYNEAEPVYFRNQAVAPKQYAYDALYQ 361  
Qy 363 LISATGREMANIGQSHQFP--SPALPSDNNTYNTYTRYTDRGNLTKIQHSSPATON 420

Db 362 LVSSGSSDALRQOTSLPLITP-IPLDSSQYVYAEKYSYDQAGNLKLSHNG---AS 417  
Qy 421 NYTNTITVSNRNSRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480  
Db 418 QYTTNVYVDSNRGIMWRQ-GEDIPDIAASFDRAGNOQALFPGPRLWDTRNQLSRVHMV 476  
Qy 481 KRDGANDREWRYSGDGRMLKINEQOASNAQTORVTVLPNLELRLTONSTATTEDL 540  
Db 477 VRE-GGDNDWEGYLYDSSGMRIVKRSTRKTQTTQTTDTTLLPGLLELRIRQTGDRVTEAL 535  
Qy 541 QVITVGEAGRAQVRVLHWESG-KPEDIDNNQLRYSYDNLIGSSOLELDSBQOIISEBEYY 599  
Db 536 QVITVDE-GAQVRVLHWEDGTEPGGIANDQYRYSLNDHLTSSLLEVDGQOIIISKEEFY 594  
Qy 600 PYGCTALWAARNQTEASVKTIRYSGKERDATGLYYGYRYVYQPIWIGRWLSSDPAQTIDGL 659  
Db 595 PYGCTALWTARSEVEASYKTIRYSGKERDATGLYYGHRYYMPWMLGRWLNPDPAGMVDGL 654  
Qy 660 NLYRMVRNNPVTLLDPPGLMP-TIA-----ERIAALK---KNKVTDSAPS 700  
Db 655 NLYRMVRNNPGLMDPNGNAPINVADYSFVHGDLVYGLSKERGRYLKLFNPNFNMKSIDS 714  
Qy 701 PANA-----TNVAINI-----RPP--VAPKPSLPKASTSQPTTHP-----I 735  
Db 715 PAMVIDQYNNVALSITNQYKVEELMKFQDPQKAARKIKVPEGNLSRNNENYPLWHDYI 774  
Qy 736 GAANIKPTSGSSIVAPL-----SPVGNKSTSEISLSPESAQSSSS 776  
Db 775 NIGEAKAFAKASHIFQEVKNGYKDYHYHKLDRMIESPLWKRGSKLGLBIAA----- 828  
Qy 777 TTSTNLQKK-SFTLYRADNRSEFQSKFPBGKAWT 812  
Db 829 ---TNQRTKIHFVL---DNLNIEQVTVKGGSGQSIT 859

## RESULT 15

US-10-754-115-42  
; Sequence 42, Application US/10754115  
; Publication No. US20040208907A1  
; GENERAL INFORMATION:  
; APPLICANT: Hey, Timothy  
; APPLICANT: Schleper, Amanda  
; APPLICANT: Bevan, Scott  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Li, Ze Sheng  
; APPLICANT: Ni, Weiting  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Don  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Meade, Thomas  
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
; FILE REFERENCE: DAS-104XC1  
; CURRENT APPLICATION NUMBER: US/10/754,115  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR APPLICATION NUMBER: US 60/441,723  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 930  
; TYPE: PRT  
; ORGANISM: Paenibacillus strain DAS1529  
US-10-754-115-42

Query Match 31.6%; Score 1581.5; DB 4; Length 930;  
Best Local Similarity 42.2%; Pred. No. 1.5e-93;  
Matches 370; Conservative 128; Mismatches 284; Indels 95; Gaps 23;  
Qy 8 LYQKTPVSVYDNRGLIIRNIDFHRITANGDPDTRITHQYDIHGLNQSIDPRLYEAKQ 67  
Db 6 IYRGTPITSVVDNRNLRIQLQYNRAAEDPADECILENTYTPLSYLGSSMDPRLFSQYQ 65

Search completed: February 16, 2006, 21:54:20  
Job time : 90.0796 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 21.6251 Seconds  
(without alignments)  
3670.210 Million cell updates/sec

Title: US-10-754-115-47  
Perfect score: 5005  
Sequence: 1 MKNIDPKLYQKTPVSYVDN.....DAEISFLTTIPLKNVPHKR 960

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/iaa/5 COMB.pap.\*
- 2: /cgn2\_6/ptodata/1/iaa/6 COMB.pap.\*
- 3: /cgn2\_6/ptodata/1/iaa/H COMB.pap.\*
- 4: /cgn2\_6/ptodata/1/iaa/pCTUS COMB.pap.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE COMB.pap.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4545	90.8	954	2	US-09-251-645-12
2	2740	54.7	1043	2	US-08-851-567B-61
3	2537.5	50.7	915	2	US-09-817-514A-6
4	329	6.6	1627	2	US-09-328-352-6604
5	322.5	6.4	998	2	US-09-253-991A-28424
6	303.5	6.1	1439	2	US-09-543-681A-7560
7	300.5	6.0	1426	2	US-09-492-709A-340
8	291	5.8	1586	2	US-09-543-681A-5329
9	291	5.8	1626	2	US-09-252-991A-23805
10	288.5	5.8	1596	2	US-09-328-352-5542
11	287.5	5.7	974	2	US-09-253-991A-23640
12	277	5.5	1377	2	US-09-711-164-467
13	261.5	5.2	1665	2	US-09-543-681A-4476
14	245	4.9	1584	2	US-09-251-645-6
15	193.5	3.9	1128	2	US-09-252-991A-11032
16	186	3.7	804	2	US-09-328-352-5545
17	179	3.6	2736	2	US-09-253-991A-30227
18	178.5	3.6	10182	2	US-09-134-001C-3159
19	178	3.6	408	2	US-09-543-681A-5372
20	170	3.4	344	2	US-09-328-352-6657
21	167.5	3.3	1016	2	US-09-134-000C-6204
22	167.5	3.3	1306	2	US-09-538-092-330
23	165.5	3.3	2137	2	US-09-134-001C-4463
24	164.5	3.3	5024	2	US-09-710-279-2964
25	164	3.3	2283	2	US-10-172-502-4
26	162.5	3.2	1394	2	US-08-296-791-2
27	162.5	3.2	1394	2	US-09-839-996-2

28	162.5	3.2	1394	2	US-10-080-505-2	Sequence 2, Appli
29	162.5	3.2	1394	2	US-10-645-655-2	Sequence 2, Appli
30	162.5	3.2	1394	4	PCT-US95-10661A-2	Sequence 2, Appli
31	161	3.2	1395	2	US-10-080-505-7	Sequence 7, Appli
32	160.5	3.2	3290	2	US-09-328-352-5486	Sequence 5486, Ap
33	159	3.2	1187	2	US-09-949-016-6513	Sequence 6513, Ap
34	159	3.2	2169	2	US-09-949-016-6930	Sequence 6930, Ap
35	158	3.2	1537	1	US-08-325-267A-2	Sequence 2, Appli
36	157	3.1	1335	2	US-09-134-001C-3716	Sequence 3716, Ap
37	157	3.1	1848	2	US-08-296-791-6	Sequence 6, Appli
38	157	3.1	1848	2	US-09-839-996-6	Sequence 6, Appli
39	157	3.1	1848	2	US-10-080-505-6	Sequence 6, Appli
40	157	3.1	1848	2	US-10-645-655-6	Sequence 6, Appli
41	157	3.1	1848	4	PCT-US95-10661A-6	Sequence 6, Appli
42	157	3.1	3696	2	US-09-134-001C-5080	Sequence 5080, Ap
43	156.5	3.1	1439	2	US-09-134-000C-6133	Sequence 6133, Ap
44	156.5	3.1	1589	2	US-09-543-681A-4998	Sequence 4998, Ap
45	155	3.1	2353	2	US-09-377-155-33	Sequence 33, Appli

## ALIGNMENTS

### RESULT 1

US-09-251-645-12  
; Sequence 12, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251.645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 954  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-12

Query Match	90.8%	Score	4545	DB	2	Length	954
Best Local Similarity	90.4%	Pred.	No. 0				
Matches	868	Conservative	40	Mismatches	46	Indels	6
Gaps	2						
QY	1	MKNIDPKLYQKTPVSYVDNRLGIIRNIDFRTTANGDPDTRITRHOYDIHGHLSQSDP	60				
Db	1	MKNIDPKLYQKTPVSYVDNRLGIIRNIDFRTTANGDPDTRITRHOYDIHGHLSQSDP	60				
QY	61	RLYEAKQNTNTIKFNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR	120				
Db	61	RLYEAKQNTNTIKFNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR	115				
QY	121	QYETSSLPGRLLSVAEQPESKTSRITERLLIWAGNTEAKDHNLAGQCVRHYDTAGVTRL	180				
Db	116	QYETSSLPGRLLSVAEQPESKTSRITERLLIWAGNTEAKDHNLAGQCVRHYDTAGVTRL	174				
QY	181	ESLSITGTVLSQSSQLLIDTQEAANWTDNETVWQNLADDIYTTLTSTFDATGALLTQTD	240				
Db	175	ESLSITGTVLSQSSQLLIDTQEAANWTDNETVWQNLADDIYTTLTSTFDATGALLTQTD	234				
QY	241	KGNIQRLAYDVAGQLNGSWLTKGQTEQVVIKSLTSAAGQKLREHGNVDITEYSBP	300				
Db	235	KGNIQRLAYDVAGQLNGSWLTKGQTEQVVIKSLTSAAGQKLREHGNVDITEYSBP	294				
QY	301	TQRLIGIKTRPSDTKVQLQDLRYEYDVPVGNVISINDAEATRFWHNKVMPENTYD	360				

[illegible]

Qy	1	MKNIDPKLYQKTPVSVVDNRGLLIRINIDFRHTTANGDPDTRIIRHQQYVDIEHGLNQSDP	60
Db	1	MSPSETLTYTQTPVSVLDNRGLSIRIDIGFRIIRVIGGGDTRVTRHQVDARGLHNLSDP	60
Qy	61	RLYBAKOTNTNITKENFLWQVLDLTCGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR	120
Db	61	RLYDAKQADNSVKFNFWQHDLAGHALRTESVDAGRTVALNDIEGRSVMTNATGVRQTR	120
Qy	121	QYETSISLPGRLLSVAEQPPEKTSRIIRTERLIWAGNTEAEKHNLNLAGQCVRHYDTRAGVTRL	180
Db	121	RYEGNTLPGRLLSVSEQVFNOESAKAVTERFIWAGNTTSEKEYNLISGLCIRHYDTRAGVTRL	180
Qy	181	ESLSLTGTVLSOSSOLLIDTOEANWTKGNETVWQNLADDTITTLTSTFDATGALLTQTD	240
Db	181	MSQSLAGAMLSQSHOLLAEGQANWGGDETFWQGLMASEVYTTQSTTNALGALLTQTD	240
Qy	241	KGNLQRLAYDVAGQLNGSWLTKGQTEQVIITKSLTYSAAQOKLREEHGNVITEYSIEPE	300
Db	241	KGNLQRLAYDIAGQLKGSWLTVKQSEQVIYKLSLSWSAAGHKLREEHGNVITEYSIEPE	300
Qy	301	TQRLIGIKTRR - - - PSDTKVLQDLRYEYDVPGVNISIRNDAEATRFWHNKQMPENTYT	356
Db	301	TQRLIGITTRRAEGSQSGARVLQDLRYKYDVPGVNISIHNDAAEATRFWRNKQVEPENRYV	360
Qy	357	YDSLXYQLISATGREMANIGQOSHQPSPAL - - PSDNNTYTNVTRYTYDRCGNLTKIOHS	414
Db	361	YDSLXYQLMSATGREMANIGQOSNQSPSVIPVPTDDSTYTNVTRYTYDRCGNLQVIRHS	420
Qy	415	SPATQNNYTNITVSNRSNRVLSTLTEDPAQVDALFDAGGHQNTLISGONLNNWNRGEL	474

```

Db 421 SPATQNSYTTDITVSSRSRAVLSTLTDPTRVDALFDSGGHOKMLIPQNLDMNIRGEL 480
Qy 475 QOVLTVKRDKGANDREWYRSGDGRMLKINEQASNNAAQORVYVLPNLELRLTQNST 534
Db 481 QVTPVSRN--SSDSEWYRSGDGRMLKINEQASNNAAQORVYVLPNLELRLTQNST 538
Qy 535 ATTEDQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIS 594
Db 539 KTTEDQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIS 598
Qy 595 EBEYYPYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAG 654
Db 599 QBEYYPYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAG 658
Qy 655 TIDGLNLRYMVRNPNVTLDDPGLMPT 681
Db 659 TVDGLNLRYMVRNPNVTLDDHGLAPS 685

```

RESULT 3

```

US-09-817-514A-6
; Sequence 6, Application US/09817514A
; Patent No. 6639129
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-6

```

```

Query Match 50.7%; Score 2537.5; DB 2; Length 915;
Best Local Similarity 70.1%; Pred. No. 2.8e-182;
Matches 488; Conservative 75; Mismatches 108; Indels 25; Gaps 6;

Qy 1 MNIDPKLYOKTPTVSVDNRGLIIRNIDFHTTANGDPDPTRIHQYDIHGLNQSIDP 60
Db 1 MSSYNSAIDOKTPTSKVLNDRKLNVRTLEYLRTOADENSDELITFYEFNIFGFQVKSTDP 60
Qy 61 RLYEAKQTNNTIKPNFLQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120
Db 61 R-----KNNQSGPNFIRVFNLAGQVLRBSVDAGRTITLNDIESRPVLIINATGVRQNH 115
Qy 121 QYETSSLPGRLLSVAEQTPE-EKTSRITERLIWAGNTEAKDHNLAGQVRHYDTAGVTR 179
Db 116 RYEDNTLPGRLLAITEQVQAGEKT--TERLIWAGNTPQEKDYNLAGQVRHYDTAGLTQ 172
Qy 180 LESLTLGTVLSQSQQLLIDTOEANWTGNETWQWMLADDIYTLTSTFDATGALLTQTD 239
Db 173 LNSLSLAGVLSQSQQLLTDQADWTGDEDSLMQOKLSSDYITQSNNTDAGALLTQTD 232
Qy 240 AKGNITQRLAYDVAGLQNSWLTAKGTQVITKSLTYSAAQGLREHGNVDIETYSYEP 299
Db 233 AKGNITQRLAYDVAGLQNSWLTAKGTQVITKSLTYSAAQGLREHGNVDIETYSYEP 292
Qy 300 ETQRLIGIKTRRPSDKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKVPENTYTYDS 359
Db 293 ETQRLIGIKTRRPSDKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKVPENTYTYDS 352
Qy 360 LYQLISATGREMANIGQSHQFPSPALPDSNNTYNTYTRSYSDHSGNLTQIRHSSPATQ 419

```

```

Db 353 LYQLISATGREMANIGQSHQFPSPALPDSNNTYNTYTRSYSDHSGNLTQIRHSSPATQ 412
Qy 420 NNYTNTITVSNRRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWRGELQOQVTL 479
Db 413 NNYTVAITLNSNRNGVLSLTLPDQVDTLFDAGGHQNTLISGQNLNWRGELQOQVTL 470
Qy 480 VKRDKGANDREWYRSGDGRMLKINEQASNNAAQORVYVLPNLELRLTQNSTATTED 539
Db 471 ---NNGPCN--EWYRYSNGMRQLKVSQPTQNTTQQQRVYVLPGLLELRTTQSNATTTEE 525
Qy 540 LQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYV 599
Db 536 LHVITLGEAGRAQVRVLHWSGKPEDVNNQLRYSYDNLIGSSQLELDSGQIISSEYV 585
Qy 600 PYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAGTIDGL 659
Db 586 PFGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAGTIDGL 645
Qy 660 NLYMVRNPNVTLDDPGLMPTIAERIAALKKNKVT 695
Db 646 NLYMVRNPNVTLDDPGLMPTIAERIAALKKNKVT 672

```

RESULT 4

```

US-09-328-352-6604
; Sequence 6604, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6604
; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6604

```

```

Query Match 6.6%; Score 329; DB 2; Length 1627;
Best Local Similarity 22.2%; Pred. No. 2e-15;
Matches 197; Conservative 105; Mismatches 293; Indels 294; Gaps 40;

Qy 14 TVSYVDNRGLIIRNIDFHTTANGDPDPTRIHQYDIHGLNQSIDP 54
Db 696 TSYYTDERGNVL-----TTQDDGATSYFEYDEKQNLTCGMVDAEQGRWPKQYDGSNL 748
Qy 55 NQSIDPRLYEAKQTNNTIKPNFLQY-DLTGNPLCTESIDAGRTVTLNDIEGRPLLTQ 113
Db 749 IKEIDPLKHETAYVYNAM--GLVTSITDAKGGSKSLKYDDQGNLISYTDGSGK----- 799
Qy 114 TGVQTR-QYETSSLPGRLLSVAEQTPEKTSRI-----AEQTPBEKTSRI-----TERL 150
Db 800 ---ETKWQYDER---GRVISIENALNQKVEFYTELTLLENRPFIKGLPLNAGOLEKI 852
Qy 151 IWAGNTEAKDHNLAGQVRHYD-TAGVTRLESLSLTGTVLSQSQSL----- 196
Db 853 KHADGTEHFHDAEGRLLAHVDPKQNTRYE-YDEAGLILSRDLDALNHLKLYKWDRLGR 911
Qy 197 ---LIDTQEANWT-----GDNETVWQ-NMLADDIYTLTSTFDATGALLTQ 237
Db 912 LTRLINENGASYQFFDVASRLVKIEDPDGKETVYVHDEKSGQLATSEVASAYQDQDK 971
Qy 238 TDA-KGNIQRLAYDVAGQLN-----GSLWTLKGTEQV-----IKSLT----- 275
Db 972 RAAPKDRITQIFDSMGRLEQRTAGYGHYGLEBEK-QTEEFAYDYMGRIIQAQNAQSNL 1030
Qy 276 ---YSAAGQKLRHEHGN-----VITEYSVEPETQRLIGIKTRRPSDKVLQDLRYEYD 326
Db 1031 QWIFYDAAG-NLVQHEHQDYKINKTAVWKHQYDEINDR---IKTRP-DGQVIDWLTGSG 1085

```





```
Db 1248 NPLQIDPGLLQT-----WDDARSGACTGVC----- 1275
Qy 728 SQPTHTPGANIKPTTSGSSIVAPLSPGVGNKSTSEISLPESAQSS-----SS 775
Db 1276 -----GVLRIIGSKFSDADALDA--KETQNRSLCNDMEYSGIVCKDTNGKYPAS 1327
Qy 776 STTSTNLOKKSFTLYR 791
Db 1328 KAETDNLKRESYPLKR 1343

RESULT 8
US-09-543-681A-5329
; Sequence 5329, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5329
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; NAME/KEY: UNSURE
; LOCATION: (1576)
; FEATURE:
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-5329

Query Match 5.8%; Score 291; DB 2; Length 1586;
Best Local Similarity 21.5%; Pred. No. 1.4e-12;
Matches 209; Conservative 121; Mismatches 358; Indels 282; Gaps 43;
Qy 18 YDNRGLIIR-----NIDFHTTANGDPDTTRITHQYDIHGLNQSIDPRLYE 64
Db 704 YDERGNLTQWTDPEQOSTCYDWDKDFAPPTAQTLPNGAAWHWEYNEHGDIRRVDPGLHI 763
Qy 65 AKQTNNTKPNFLWQYDLTGPNLCSTESIDA-----GRVTVLNDIEGRP-LITV 111
Db 764 TR-----LAWD-----DQGLCLGQVDKGNETHYRNARGQLIEQDCSGYPTTLTY 810
Qy 112 TATGVIQT---RQVETS---SLPGRLLSVAEQTPPEKTSR-----ITERLIWAGNTEAEK 160
Db 811 DDWQQLRSLTWAQNETTYTTFSEAGLLLT--ECLPDGTENYDYDATGQLV--GITDAGE 866
Qy 161 DH-----NLAQGVV-----HYDTAGVTRLESLSLTGTVLSSQLLIDTOEAN 204
Db 867 RHILLRRNRGQVIARRDPAGHWHLFHYDTFG--RMOAL----- 903
Qy 205 WTGNETWQNLADDIYTTISTPDATGALLTQTDKAGNIQRLAYDVAGOLGNSWLTLKG 264
Db 904 ---ENEQGEQYRF-----EYDALHRLTDEHDLIGQQKHQYQVDMGNVTQIKTTPGP 951
Qy 265 QTEQVI-----IKSLTYSAAQKLRHEHGNVITEYSYEPETQRLIGI-----KTR 310
Db 952 SIDTPIPLSPQVTFYGVKQVRLIFRENA-DYRTEYLYQPLSVTLRRVPMVWHEAERTG 1010
Qy 311 RPSDTKVLQDLRYEYDPVGNVI-----SIRND----- 337
Db 1011 TTARVEYQDALTFYTKVQGVLRBASARGDYQHHDYVLNLTTRTELPHORAFEYLYYSG 1070
Qy 338 -AEATRFHWNQKWPENTYTTDSLYQLISAT-----GREMANIGQQ--SHOF 381
Db 1071 HLQQTQWRDNEQLTVLAEYQYDRDLRLTSLTSGALDNETGYDCRGRTHTHVARQWNASQF 1130
Qy 382 PSPALP-----SDNNTVTN---YTRTYVDRGGNLTQIQHSSPATQ 419
```

```
Db 1131 VTPVIDRRYRWDKRNQLIERSVSYQTGEVFTAGHWYVHSYQYDPLGQLTAHLGSGVQTEH 1190
Qy 420 NNYTNTITVNSRNRVAVLSTLTEDPAQVDALFDAGG-----HONTLISGONLWNWTRGEL 474
Db 1191 FLYDAAANLLTRPTEAPHNQVQSGDKFDYRYDGFGRMVSRYERKSGSSGQRYHYDSD--- 1247
Qy 475 QQVTLVKRDKG-ANDREWRYSGDGRMLKINQQOASNNNAQTQRTVYLPNLELRLTQNS 533
Db 1248 HRIITAVDIDQPLGYQRAEYRYDILGRRIEK-RLWKLSAIANT--VYVQH----- 1295
Qy 534 TATTEDLQVITVGEAGRAQV-----RVLHW---ESGKP-----EDIDN-----NOL 571
Db 1296 ---EPDEVYTFGWGMELVSEHSSAAPHTTVYHAYNDQSYTPLARIECTDNPLNPORAI 1351
Qy 572 RYSYDNLIGSQLELDSGQIISEEEYYPYCGTALWAARNOQTEASYKTIRYSGKERD-AT 630
Db 1352 YYTHSSLSGLPEALTNSEGEIVWQGVSAWHLQORQTRPTSTFNREQLRFQGGYFDKET 1411
Qy 631 GLTYGVYRYQPMIGRWLSSDPAGTIDGLNLYRMVRNPNVTLDDPDGLMPTIAERIAALK 690
Db 1412 GLHNTFRYAPDLGRFTQDDPIGLAGINLIYAYA-PNPLTWDPWGW--SCGSDAKILR 1468
Qy 691 KKKVYDSAPSANATNVAINIRPPVAPKPSLPKASTSSQPTTHPIGAANIKPTTSGSSIV 750
Db 1469 ENMIERGIIIP-NYKNSA-----HHIIMSNAXDVRM-RWLR 1502
Qy 751 APLSPVG---NKSTSEISLPESAQSSSTSTTNLQKKSFT-LYR-----ADNRS 796
Db 1503 RKNRLGIDINDAINGVLTSSKIKIDVGSTTIAHSKIHTALYKQNVDFRLKNIKDATE 1562
Qy 797 FEEMQSKFPE 806
Db 1563 FROELINKIAE 1572

RESULT 9
US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
; LENGTH: 1626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23805

Query Match 5.8%; Score 291; DB 2; Length 1626;
Best Local Similarity 21.2%; Pred. No. 1.5e-12;
Matches 167; Conservative 95; Mismatches 284; Indels 242; Gaps 32;
Qy 10 QKTTVSV-----YDNRGLIIRNIDFHTTANGDPDPTTRITHQYDIHGLNQSIDP-----R 61
Db 863 EQLPDGVSVRYRYDTLG-----ROITRQDESGAITRFQWDAAGRLSQITLPGGASR 913
Qy 62 LYEAQKQTNNTIKPNFLWQYDLTGPNLCSTESIDAGRTVTLNDIEGRPLLT--VTATGVIQT 119
Db 914 MY-----RYNAYGK-VTSECDEQGRMTRYEYVLDLHLVSRREINPDGSQLR 957
Qy 120 RQYETSSLPGRLLSVAQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVTH-YDTAGVT 178
Db 958 YRYENARL---LLSEIENERGER-----YRLDYHGNGLISQETGDFGRRTAYRYDLKG-- 1007
```

Qy 179 RLESLSLGTSLVSSQLLIDTQEANWTDNETVQNMALDIYTLSTFDATGALLPQT 238  
Db 1008 -----QLEKTEYDGDGSELRTTYQ-----DSTGRLLAKT 1038  
Qy 239 DAKGNIQRLAYDVAGOL-----NGSW-LTLKGTQTEVVIKS-----LTYSAAGOKL-- 283  
Db 1039 LPDGNRVDRYDTLGRVAVDDGTWPLAYEYDLRLVREHGQWATLHYADALGQLIHC 1098  
Qy 284 -----REHGNVDVITEYVEPOTRLI 305  
Db 1099 RLPDGNRVDRYQGTGLSALDNGQSLTRHQFGSRRERQOQGBELLQOYHYDEGRLLA 1158  
Qy 306 GIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWMHKKVMPENTYDLSYQLIS 365  
Db 1159 HQVSQR---QRHLQRYQRYDASGNLAALIEDSRKIGSFH-----YDPLDRLLG 1204  
Qy 366 ATGR-----EMANICQOSHOPSPALP-SDNNYTYNTRYTYTYDRGNLTKIQHSSP 416  
Db 1205 VRGETPESFVHDPAGNLLAQGGQFDARQMEVRGNRLLTQGRHFDYDAHGNLVRRRGTG 1264  
Qy 417 ---ATQNNYTNITVSNRSRAVSLTSTEDPAQVDALFDAGHQ-NTLISGNLNNWNRG 472  
Db 1265 QKLVEYSYD-----COHRLIGVSLPDGQOVVRYDAPGRRIAKQVDGRNTEFLWL 1316  
Qy 473 ELQOVTLVKRDKGANDREW---YRYSGDRMLKINEQOASNAQOTQRTVYLPNLELRLT 530  
Db 1317 E-----RLLAESGDRHYRTLYEPDSEPRLLD----- 1345  
Qy 531 QNSTATTEDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSOLELSD 590  
Db 1346 -----GEG-----PEQVE--PCYQLDHL-GTPQELTRADG 1373  
Qy 591 QIISEEYYPYGGTALWAARNQTEASYKTYRYSGKERDA-TGLYVYGYRYQPMIGRWLS 649  
Db 1374 RLCSARTRAGN-----VLKLDIAEVNPLRPGQYFPAETGLHYNHRYNPFSTGRVLT 1429  
Qy 650 SDPAGTIDGLNLYRMVRNPVTLDPDGLMPTTIAERIAALKKNKVTDSAPSANATNVAI 709  
Db 1430 PDPRLAGGLNSRYV-FNPTGWVDPLGLASKNTRCPALDNN-----PRKGINSTS 1480  
Qy 710 NRPVAP 717  
Db 1481 GSKKPAGP 1488

## RESULT 10

US-09-328-352-5542  
; Sequence 5542, Application US/09328352  
; Patent No. 6562958

## GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT APPLICATION NUMBER: US/09/328,352

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5542

; LENGTH: 1596

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-5542

## Query Match

Best Local Similarity 5.8%; Score 288.5; DB 2; Length 1596;

Matches 179; Conservative 123; Mismatches 316; Indels 189; Gaps 39;

Qy 19 DNRGLIIRNIDPH-RTTANG-----DPPTRITRHOYDIHGHLNQSIDP----- 60

Db 637 DKKKRTIQIDPDGRETQEQYNDQQLVKIVQPNQGIIRFAYNKQGNLVKIDPGESIWK 696

Qy 61 RLY-EAKQTNNTIKP-NFLWQYDLTGNPLCTESIDAG---RTVTLNDIEGRPLLTVTATG 115

Db 697 REYDENRVNKBINPLGHITQYKNNQNQLVEIDAKGGVKKIQVNEI-GQMSITDCSG 755  
Qy 116 VIQTRQYETSLPRLLSVAEQTPPEKTSRI-----TERLIWAGNTEAEKQHNH 164  
Db 756 KSSITWEYDED---GAL---TABQATANNKVQYFYSTKGRDKGQLQSIIYPDGLKEFFEHE 810  
Qy 165 AGQCVRHYDTAGVTRLSLSLTGTVLSSOLLIDTQEAN-----WTGDNETVWQNMALD 219  
Db 811 EGRLLKHTDTRKGLVTEYKYNQV-----LLEQRIDANRHSVAYQWQDKQRIQKLIQ 862  
Qy 220 DIYTLSTFDATGALLTQTDAGNIQRLAYDVAGOL-----NGSMLTLKAGQTE---QVI 270  
Db 863 NOAEYLFQYNGPYGLIREQAFDGBEKHYSYNNENGRKLFQIRPNILUT---QDYIYADGQIA 919  
Qy 271 IKSLTYSAGOKLREEHG---NDVITEYSYEPET---QRLIGIKTRRPSDTKVLQ--- 319  
Db 920 SKSFTHLHTGQKQTEQFDYNLSQLSRASNEVSQIDLYRNALGQLVREHQHVKIFELKPL 979  
Qy 320 ---DRIYEDVPVGNVI-SIRNDAEATRF-----WHNQKVMPEYTYDLSYQLI 364  
Db 980 TAVLHYEYDELGNLITIRPDGHTLNHLVYSGSHIYAIGLANQEVV---SFQDDLHR-- 1034  
Qy 365 SATGREMAN-----IGQOSHOPSPALP-SDNNYTYNTRYTYDRGNLTKIQ 412  
Db 1035 -ETRLLANGLMQTKQYNDVGLSSQFNQ---EQTQDYLYQYQAHKTHYDKNYLLSQVE 1091  
Qy 413 HS-----ATQNNYTNITVSNRSRAVSLTSTEDPAQVDALFDAGHQ 458  
Db 1092 DSRGLKLYQYDPIGRLLIAQSLHKTESFNFDPAGNLIDSESVLSPAQI-----KN 1142  
Qy 459 TLI---SQONLNNWTRGELQVTLVKRDKGANDREW-----YRYSGD 498  
Db 1143 NLIKSYKGYQYDQVGNVTBI---QAGKNLKLTDNQNLIRSDNNGLVTEYGDV 1198  
Qy 499 GRMLKINEQOAS-----NNAQ---TORVTVLPNLELRLTSTATTEDLQ 541  
Db 1199 GRRLYKTKAKELTLFGWDGDLMIWESFKSAQNTYTKHYIYBPDSFVPLQ---AGYKDFIQ 1256  
Qy 542 VITVGEAGRAQVR-----VLHWESGKPEDIDNNQLRYSYDNLIGSSOLELSD 593  
Db 1257 LIETPDYQYQYKPYSIYKDPVNRNLGK-ERTALEQFTFYHCDQVGTPTMTNIRGECV 1315  
Qy 594 SEEBEYYPYGGTALWAARNQTEA-SYKTYRYSGKERD-ATGLYVYGYRYQYQWIGRWLSSD 651  
Db 1316 WEILQDTWGAVSQIKALNQDNPFQONLRFQYQYDRETELHYNRYRYEPHSARYSKD 1375  
Qy 652 PAGTIDGLNLYRMVRNPVTLDPDGL 678  
Db 1376 PIGLEGGNNTSSYV-SDPNQWIDPKGL 1401

## RESULT 11

US-09-252-991A-23640

; Sequence 23640, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 23640

; LENGTH: 974

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-23640





Db 1215 GRITQDPGLKGMNFYOYPL-NPVTNTDPLGL 1247

RESULT 13  
US-09-543-681A-4476  
; Sequence 4476, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4476  
; LENGTH: 1665  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4476

Query Match 5.2%; Score 261.5; DB 2; Length 1665;  
Best Local Similarity 21.1%; Pred. No. 2.6e-10;  
Matches 184; Conservative 124; Mismatches 348; Indels 215; Gaps 42;

Qy 31 HRTTANGDPDTRITRHOYDIHGLNQSIDP---RLYAKOTNNFK---PN-FLWQYDL 82  
Db 816 HFTSAFGE---WQSDVNGNLITVDPOQRQVYHSQHGELLKAITPAGQWQHY 871  
Qy 83 -----TGNNPL---CTE---SIDAGRTVLNDIEGRPLLTVTATGVIOTROYETSSLPGR 130  
Db 872 NPAHQIKTTPYQHSSTVHSDGLLHYTD-----ALNHTTRYQVS----- 914  
Qy 131 LLSVABQTPKTSRITERLIWAGNTEAKDHNLAGQCVRHYDTAGVTRLSLSTGVTL 190  
Db 915 -----TEHASTNGSLSKILLPDGVEQOIDDYSERRVIAVTDGEGKTRYRYGPPDVL 967  
Qy 191 SQSSOLLIDTOBANWGTNETWQNMADDIVTLLSTPDATGALLTQTDAGNIQRLAYD 250  
Db 968 ---AMIRPDGSEIRFEYSLTLKKVNVANGEVILYERDKAGIIRFVDFGTREICRYD 1024  
Qy 251 VAG-----QLANGSWTL-----KQTEQVIAIKSLTYS-AAGQKLREBHG 288  
Db 1025 RLGRRIATRYPDNHELRYNESGLVQSEWFEDEQSSRCLSTAQYSYNARQQLIKATN 1084  
Qy 289 NDVITEYSYEP-----TORLIGIKTRRPSDTKVL-----QDLRYEYDPVGNVISIR 335  
Db 1085 PDSVVEPEYDDQGRLCBRINEQEIHVQWNEADNTLALTRFGERELHYAFGALGELSLQ 1144  
Qy 336 -NDAEATRFMN---QKMPENT---YTVDLSYQLISATGREMANIGQQSHQFPSPALPSD 389  
Db 1145 VNQHAPLQFSYNAVQVEYLRRSRAGFVNSHYTATGLLAHORAGRGTEQFLOSIQAHQQ 1204  
Qy 390 NNTYNTYRTYTRDGGNLTQIHS-----SPATON----- 420  
Db 1205 PPFCTDVRHSYQDRAYNVGIEDDRWRQTRVHYNANDQITETQYSPQWQDEKFOYDN 1264  
Qy 421 -NYTNTIT-----VSNRSNRAVLSTTEDPA-QVDALFDAGGHQNTL--ISGO----- 464  
Db 1265 NLNITEHLTTPSSMVPDAQAMQLQFQQQAGVTRRYTAKGYQDYHYDVNGRLAKKI 1324  
Qy 465 -----NLNNTTGLQVTLVKRDKGANDREW-VRYSGDGRMLKINEQQAQN 512  
Db 1325 VHTGRFRERWYLNNTQNLACTPFGDC-----WHYTFAGRLSK-FKTVUSD 1376  
Qy 513 NAQTORVTLPLNLERLTQ-----NSTATTEDLOVITVGEAGRA-QVRVLHWESGKPBEDI 566  
Db 1377 LAHIDPL--FPQIKPKITWYLSGSDQLIBETPIYADGTLANAQQVQWLY---QPGEI 1430  
Qy 567 DNNQLRYSDNLIQSSQLESEGGIISBEYYPYGGTALWAAR----- 610

Db 1431 -TPTARYOQGL---HYVVDHQG---TPREIFSEGGOASWAGRLNTWQWQFWRDRDK 1483

Qy 611 -----NOTEASYKTIRYSGK-ERDATGLYYGYRYQYQWIGRWLSSDPAGTIDGLNLRYM 664  
Db 1484 AENDPNYTECPF---RFAGQYDEBESGLYNNRFRYDRYDRETQYLSDFPIGLGLGGLNPYGY 1540

Qy 665 VRNPNVTLDDPGLMPTTAERIAALKKX---KVTDSPASPANATN-VAINIRPPVA--- 716  
Db 1541 V-HCPTGWVDFGLAGDCDCLLVLATNPRSKTVQEMAEDLANRNRNTVTFSTPTSMGH 1599

Qy 717 ---PKPSLPAKSTSSQPTTH-PIGAANKP 742

Db 1600 IDLGRSHFDKVTQADIPTPHVQSPINAP 1630

RESULT 14  
US-09-251-645-6  
; Sequence 6, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1584  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-6

Query Match 4.9%; Score 245; DB 2; Length 1584;  
Best Local Similarity 19.3%; Pred. No. 4.2e-09;  
Matches 202; Conservative 126; Mismatches 307; Indels 414; Gaps 44;

Qy 14 TVSYDNRGLAIRID---FHRTTANGDPDTRITRHOYDIHGL-NQSIDPRLYEAKQTN 69  
Db 552 TVTGFDAHMSKKNVTSIYTHQLRKVDNVHVIDQSYDLLGRITGQIIDP-----GTA 605

Qy 70 NTIKPNFLWQDLTGNPLCTESIDAGRTVTNLNDIEGRPLLTVTATGVIOTROYETSSLP 129  
Db 606 REIKRNVYQY-----PGGDE-----NDF-WPVMIEVDSQGVARKTHYDGM--G 646

Qy 130 RLLSVABQTPK---TSRI---TERLIWAGNTEAKDHNLAGQCVRHYDTAGVTRLE--- 181  
Db 647 RICSIEQDDGAGTSGIYQGYTRKVL-----RQDVLGQLSKESIN 690

Qy 182 ---SLSLCTVLSQSSQLLIDTQEANWTDGNETWQNMADD----- 220  
Db 691 DMLNLSANPLVRLATPLVTTKYDG-----MGNLYSTEYSDGRIEIHDPITRTIT 745

Qy 221 -----IYTTLS-TFDTAGALLTQTDAGNTQRLAY 249  
Db 746 QGVKGLGLMLNIQNNFEOPASIKAVYPDGTIYSTRYDGFGRVTVTETDAEGHATQIGY 805

Qy 250 DV-----AGOLNGSWL-----TLKQOT 266  
Db 806 DVFDRIVVKTLPDGTILESAYASPSHEBLSALNVNQLGALVYDGLGRVISTVGRK 865

Qy 267 EQVI-----IKSLT-----YSAAGQKLREE 286  
Db 866 TEYLYGPGQDKPIQSIITPSHNKQNMVLYLGSVMSKFTTGTQDQNFYHSKGTLLSAS 925

Qy 287 HGNVITEYSYEP---TORLIGIKTRRP-SDTKVLQDUR-----YEYDPV 328

Db 926 EGVSQ-TNYSYPPGVLQREDFLRDNKPIISGEVLYTMSGLIQRKDSFGHNHYSDAQ 984  
Qy 329 GNVTSIRNDABATRFHNNQKMPENTYYSLSYOLISATGREMANIGO-----376  
Db 985 GRLVKTEQDAQYA-----TPEYDNVGRLLITTTTKDTTSLQSLVTKIEYADFDR 1032  
Qy 377 -----QSHQPPSPAL 386  
Db 1033 EIKRSLISDFSIOVITLSYTKNNISOIRITSIDGVMMKERYQYDNNORLSQYQCEGEQS 1092  
Qy 387 PSDNNTYNTYTRTYDRGKNLTKIOHSSPATNNYTT-NITVS---NRSNRAVLSTLITE 442  
Db 1093 PIDHTRVLNQOIYHYDQWGNIKEL-----DNTYRDKETVDYHFSQADPTQLIRITS 1145  
Qy 443 DPAQVDALFDAGGHONTLISQNLNWNTRGLOQVTLVKRDKGANDREWYRSGDGRM 502  
Db 1146 DKQIELSNDANGNLTRDEKGTLLIYDQNNRLVQV---KDLRGN-----1186  
Qy 503 LKINEQOASNNACQRTVTLNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWESGK 562  
Db 1187 LVCYSQYDALNKLUT-----AQVLANGTVNRQ-----HYASGK 1218  
Qy 563 PEDIDNNQLRYSYDNLICSSQLELDSEGOIIEBEYYPYG---GTALWAARNOTE---AS 616  
Db 1219 ---VTNIQLGDEAITLSSDKORIGHQSAKNGQSVYYQYGDHNSTVIASQNELMALIS 1275  
Qy 617 YKTIRY-----SKERD-ATGLYYY--GYRYQYPMIGRWLSSD---PAGTIDGLN 660  
Db 1276 YTPYGFRLSISLPLNGAQVDPVTGWYFLNGYRVFENPVLMRPHSPDSWSPPFGR-GGIN 1334  
Qy 661 LYRMVRNPNVTLDPDGLMPT-----TAERIAALKKKNVTDAPSAPANATVA 708  
Db 1335 PYTYCOGDPINRIDNGLHLSAGGILVGAIGIIVGIVSLGAGAAISAGLIAAGGALGA 1394  
Qy 709 INIRPPVAPKPSLKASTSSOPT-----HPIGAANI-----KPTTSG-----SSIV 750  
Db 1395 I-----ASTSALAVTATVIGLAADSIGIASAALSEKDPKTSGLNWLISAGL 1440  
Qy 751 APGSPVGNKSTSEISLPESAQSSSSTTS 779  
Db 1441 GVLSFGISAITFTTSSLVKARSQSQAQVA 1469

## RESULT 15

US-09-252-991A-31032  
; Sequence 31032, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31032  
; LENGTH: 1128  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31032

Query Match 3.9%; Score 193.5; DB 2; Length 1128;  
Best Local Similarity 18.3%; Pred. No. 1.8e-05;  
Matches 181; Conservative 110; Mismatches 301; Indels 395; Gaps 41;

Qy 4 IDPK-----LYQKPT---VSVYDNRGLIIRNIDF---HRTTANGDPDTRITRHQYDI 50  
Db 257 IDPQSPGLDMLVRYQFSDSGDLIAVRDLRGQVRVFAWREHMLVAHGEPGLGVREYWDV 316

Qy 51 HGHLNQSIDPRLYEAKQTNNTIKENFLWQYDLTGNPLCTESIDAGRTVTLLNDIEGRPLLT 110  
Db 317 HAP-----HGRVVVKQ 326  
Qy 111 VTATGVITQTRQY-----ETSSLPGRLLSVAOEPEEKTSRITERLIIWAGNTEAKOHNLA 165  
Db 327 IEAGGLTRTRYLRDATERVSDSLGRV--ERYEPAGEGQRRWTALVRADGSRSEFDVDF 384  
Qy 166 GQCVRHVDTAGVTRLESLSUTGVVLSOSSQ-----LLDITQBA---NWT-- 206  
Db 385 GRLVAMRDPLGRETRRRRRQGRMLEESPCKARYRKRVRDEETGLLVELEDAMQRRWTFE 444  
Qy 207 -----GNETVWQNMLADDITVTLIS-- 226  
Db 445 RDERGNATTVRGPAAGSTRYAYEDRPLDRPTRIVDPGGERRLERNRFLGLAULTDCSGQ 504  
Qy 227 ---TFDATGALLTQTDAGNIQRLAYDVAGQL-----NGSWLTLKGQTEQVIKSLTY 276  
Db 505 VMRYDYDNEGRVASSDPLGQLTERRYDPLQLIGLELADGSAL-----SVEY 552  
Qy 277 SAAGQKUR--BEHGNVITYEYEPETORLIGIKTRRPSDTKVLODIERYEYDVPQNVISI 334  
Db 553 DALGROTRIADAEGHATLFSWGH-----GDLARVSDAGG-GELSYLHDEAGRLVAL 603  
Qy 335 RNDAEATRFHNNQKMPENTYTYDSLYQLISATGREMANIGQOSHOPSPALPDNNNTYT 394  
Db 604 TNENGV-----QAQFRYDLRLVEETQFQGR---RQRYRNAADELAREDDAD 649  
Qy 395 NYTRTYTYDRGKNLTKIOHSSPATNNYTTNITVSNRSRAVLSTLTEDPAQVDALFDAG 454  
Db 650 GRETTYAYDRGLASIR--VPATE----- 672  
Qy 455 GHQNTLISQNLNKNNTREGLOQVTLVKRDKGANDREWYRSGDGRMLK-----IN 506  
Db 673 -HAPALV--BRYRWLADGLASA-----GGADCEVRYTYDEVGNLRLESQVHADGWVY 722  
Qy 507 EQQASNNA---OTOR-----VTYLPN-----LEL-----R 528  
Db 723 SVEHSHDALGVQRTSRYGDAPPVAVWLTYGPHLHGALVGAVELAFERDALHREVRDRAR 782  
Qy 529 LTQNSTATTEDLQVITVGEAGRAQVRV---LHWESGKPED-----IDNNQ---LRYSY 575  
Db 783 DGQDDALFTQERQHAPLGRQLRSLRAGGDFWQGVRYDGLQGLVGDIDNNQVPSVREY 842  
Qy 576 D---NLIGS-----SOLELSESG-----OIIEEYYPYG----- 602  
Db 843 DLGRLASRRAGAAASTYRYDAAGNRLEGVGEYAREDAQAFAENELYSRSGFSRSEVRA 902  
Qy 603 ---GTALWAA-----RNQTEASYKTYRYSKGERDATGLYYYG----- 636  
Db 903 NQAGEGPAPWAGNVRVERIAGNRYRFDALGNLIVERIGADGELRLAYDGAQRLVHLTRDYA 962  
Qy 637 -----YRYQ-----PWIG-----RWLSSDPAGTI 656  
Db 963 DGTLEARYRYDALSRRIAKVWLRDGVQVRFQWCDGDRQCAEAFARELRTTVHEPGGFV 1022  
Qy 657 DGLNLYRMVRNPNVTLID-----PDGLMPTIARIALKKNKV-----TDSAPSPANATN 706  
Db 1023 PLLRLEOACEFPDPPPELLQLRQAFAAEG-QFLPACQVCPALGEARIAFTHTDLHGLTFLQUSD 1081  
Qy 707 VAINIRPPVAP---KPSLPKASTSSQ 730  
Db 1082 ERGLRWQGVQVDDWRVAPERQGAQP 1108

Search completed: February 16, 2006, 21:46:40  
Job time : 27.6251 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 146.155 Seconds  
(without alignments)  
4431.220 Million cell updates/sec

Title: US-10-754-115-45

Perfect score: 7901

Sequence: 1 MQNSQDFSLTSLPLKGGGA.....WFTVNEDENDTAAEYKVKVM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq 21:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	7901	100.0	1474	8	ADP18622	Adp18622 Photorhab
2	7901	100.0	1474	8	ADR21525	Adr21525 Photorhab
3	7285.5	92.2	1476	6	ABM70222	Abm70222 Photorhab
4	6065.5	76.8	1477	6	ABM70230	Abm70230 Photorhab
5	6042	76.5	1481	2	AAy33730	AAy33730 Photorhab
6	6031.5	76.3	1476	5	ABG32652	Abg32652 P. lumine
7	6025.5	76.3	1476	5	ADR21576	Adr21576 Photorhab
8	4627.5	58.6	1486	6	ABM70526	Abm70526 Photorhab
9	4595.5	58.2	1485	8	ADR21578	Adr21578 Photorhab
10	4581.5	58.0	1485	2	AAW56546	Aaw56546 Toxin tca
11	4580.5	58.0	1485	2	AAW17887	Aaw17887 Photorhab
12	4261	53.9	1506	8	ADR21529	Adr21529 Xenorhabd
13	4261	53.9	1506	9	AEb47812	Aeb47812 Native Xp
14	3971	50.3	1428	8	AAy97695	AAy97695 SepB prot
15	3971	50.3	1428	8	ADR21540	Adr21540 Serratia
16	3834	48.5	1493	8	ADR20369	Adr20369 Recombina
17	3834	48.5	1493	8	ADR21498	Adr21498 Xenorhabd
18	3106.5	39.3	1444	8	ADM61347	Adm61347 Paenibaci
19	3106.5	39.3	1444	8	ADR21520	Adr21520 Paenibaci
20	1744.5	22.1	662	3	AAy95685	AAy95685 Cosmid cH
21	1262	16.0	697	8	ADM61375	Adm61375 Paenibaci
22	839	10.6	591	2	AAr23006	AAr23006 Protein t
23	199	2.5	2060	5	AAE20967	AAE20967 Staphyloc
24	172.5	2.2	1404	7	ADC00960	ADC00960 Enterohae

25	170	2.2	593	6	AAO19846	Aao19846 P aerugin
26	169.5	2.1	14130	9	AEA07438	Aea07438 M. ulcera
27	162.5	2.1	16990	9	AEA07436	Aea07436 M. ulcera
28	161	2.0	1863	7	ADD46989	Add46989 Rat Prote
29	161	2.0	1863	9	ADX26456	Adx26456 Novel cel
30	160.5	2.0	1385	6	ABU40318	Abu40318 Protein e
31	160.5	2.0	1426	3	AAAB15983	Aab15983 E. coli p
32	160.5	2.0	1426	6	ABU14693	Abu14693 Protein e
33	159.5	2.0	1329	4	ABG28402	Abg28402 Novel hum
34	159.5	2.0	1329	4	ABG25780	Abg25780 Novel hum
35	159.5	2.0	1329	7	ADC33172	Adc33172 Human nov
36	155.5	2.0	658	7	ABO76164	AbO76164 Pseudomon
37	155.5	2.0	1400	7	ADC01365	AdC01365 Enterohae
38	155.5	2.0	2315	7	ADF05149	AdF05149 Bacterial
39	154	1.9	4630	3	AAy77177	Aay77177 S. venezu
40	154	1.9	5215	8	ADL91930	Adl91930 Streptomy
41	153.5	1.9	1028	7	ADF06896	Adf06896 Bacterial
42	152.5	1.9	1669	8	ADU07870	Adu07870 Amino aci
43	152	1.9	4630	2	AAW19629	Aaw19629 Streptomy
44	151.5	1.9	1904	5	ABB57100	Abb57100 Mouse isc
45	151.5	1.9	1904	9	ADX26385	Adx26385 Novel cel

#### ALIGNMENTS

##### RESULT 1

ADP18622	ID	ADP18622 standard; protein; 1474 AA.
XX	XX	
AC	ADP18622;	
DT	12-AUG-2004 (first entry)	
XX	XX	
DE	Photorhabdus luminescens W-14 tcdB2 protein SEQ ID NO:10.	
XX	XX	tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;
KW	KW	orally active insect toxin; insect toxin; tcdB2.
XX	XX	
OS	Photorhabdus luminescens.	
XX	XX	
PN	WO200404217-A2.	
XX	XX	
PD	27-MAY-2004.	
XX	XX	
PF	12-NOV-2003; 2003WO-IB005553.	
XX	XX	
PR	12-NOV-2002; 2002US-0425672P.	
XX	XX	
PA	(UYBA-) UNIV BATH.	
XX	XX	
PI	Ffrench-Constant RH, Waterfield NR;	
XX	XX	
DR	WPI; 2004-411735/38.	
XX	XX	N-PSDB; ADP18621.
DR	XX	
PT	XX	New isolated Photorhabdus luminescens nucleic acids, useful for
PT	XX	expressing orally active insect toxin or for generating transgenic plants
XX	XX	with enhanced resistance to insects.
XX	XX	Claim 1; SEQ ID NO 10; 118pp; English.
XX	XX	
CC	XX	The present invention describes DNA sequences from the tcd genomic region
CC	XX	of Photorhabdus luminescens W-14. Also described: (1) a transgenic
CC	XX	monocot or dicot cell having a genome comprising a nucleic acid sequence
CC	XX	that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant
CC	XX	with a genome comprising a nucleic acid sequence that encodes SEQ ID
CC	XX	NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a
CC	XX	progeny of seed; (5) a method of producing Toxin A of Photorhabdus
CC	XX	luminescens W-14 in a heterologous host; and (6) a method of producing an
CC	XX	orally active insect toxin. The nucleotide sequences are useful for
CC	XX	heterologous expression of orally active insect toxin. They can also be
CC	XX	used for generating transgenic plants with enhanced resistance to

CC insects. The present sequence represents Phototribus luminescens W-14  
 CC tcdB2, which is used in the exemplification of the present invention.  
 XX  
 SQ Sequence 1474 AA;

Query Match 100.0%; Score 7901; DB 8; Length 1474;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1474; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSQDFSIETSLPKGGGATGGEALPTGPGMAALSPLPISAGRGYAPFTLNYN 60  
 DB 1 MNSQDFSIETSLPKGGGATGGEALPTGPGMAALSPLPISAGRGYAPFTLNYN 60

QY 61 SGAGNSPGLGWCNMTIRRTTHFGVPHYDSTTFLGPEGEVLVADQPDDESLQGIN 120  
 DB 61 SGAGNSPGLGWCNMTIRRTTHFGVPHYDSTTFLGPEGEVLVADQPDDESLQGIN 120

QY 121 LGATFTVTGYSRLESFSLRLEYQPKTTGTDWLIYSPDQVHLLGKSPQARISNPSQ 180  
 DB 121 LGATFTVTGYSRLESFSLRLEYQPKTTGTDWLIYSPDQVHLLGKSPQARISNPSQ 180

QY 181 TTQTAQWLLASVSRGSIYYVRAEDDTGCEADEITHLQATAQRYLHIVVYGNRTAS 240  
 DB 181 TTQTAQWLLASVSRGSIYYVRAEDDTGCEADEITHLQATAQRYLHIVVYGNRTAS 240

QY 241 ETLPGLDGSAPSQADWLPYLDYGERSNLKTTPASTTGSWLCRODRFSRYGFEIR 300  
 DB 241 ETLPGLDGSAPSQADWLPYLDYGERSNLKTTPASTTGSWLCRODRFSRYGFEIR 300

QY 301 TRRLCRQVLYHHLQALDSKITEHNGPTLVSRLLIINYDESAIATLVFVRVGHQDGNV 360  
 DB 301 TRRLCRQVLYHHLQALDSKITEHNGPTLVSRLLIINYDESAIATLVFVRVGHQDGNV 360

QY 361 VTLPPLLEAYQDFSPRHHAHQMDVLANFNAIQRWQLVLDLKGELPGLLYQDKGAWYR 420  
 DB 361 VTLPPLLEAYQDFSPRHHAHQMDVLANFNAIQRWQLVLDLKGELPGLLYQDKGAWYR 420

QY 421 SAQRLEIGSDAVTWKQPLSVLPLOSASLVDINGDGLDWMVTGPGLRGHSORPD 480  
 DB 421 SAQRLEIGSDAVTWKQPLSVLPLOSASLVDINGDGLDWMVTGPGLRGHSORPD 480

QY 481 GSWTRFTPLNALPVEYTHPRQAQADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQVQ 540  
 DB 481 GSWTRFTPLNALPVEYTHPRQAQADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQVQ 540

QY 541 SGTITLPPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600  
 DB 541 SGTITLPPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600

QY 601 PATFENPAQVYLDLDSGPTDLIVVHTNRDLIFLKSNGFGFAPVTLRPFEGRLPDHTC 660  
 DB 601 PATFENPAQVYLDLDSGPTDLIVVHTNRDLIFLKSNGFGFAPVTLRPFEGRLPDHTC 660

QY 661 QLQWADVOGLGVASLIISVPHMSPHHWRCDLTNMPKLLNEMNNMNGVHHTLYRSSSQF 720  
 DB 661 QLQWADVOGLGVASLIISVPHMSPHHWRCDLTNMPKLLNEMNNMNGVHHTLYRSSSQF 720

QY 721 WLDEKAAALTTGQTPVCVLPPIHTLMOTETEDEISGNKLVTTILRYARGAWDGRERFRG 780  
 DB 721 WLDEKAAALTTGQTPVCVLPPIHTLMOTETEDEISGNKLVTTILRYARGAWDGRERFRG 780

QY 781 FGYYEOTDSHQLAQNAPERTPPALTKNWYATGLPVDINALSTEVWRDDQAFAGSPRFT 840  
 DB 781 FGYYEOTDSHQLAQNAPERTPPALTKNWYATGLPVDINALSTEVWRDDQAFAGSPRFT 840

QY 841 TWQNDKQVPLTPEDDNSRYFNRLKGLLSELYGLDDSTNKVPYTVTFEFSQVRLQ 900  
 DB 841 TWQNDKQVPLTPEDDNSRYFNRLKGLLSELYGLDDSTNKVPYTVTFEFSQVRLQ 900

QY 901 HTDSRYPVLSWVESRYNHYERTASDPQCSQNTILSSDRFGQPLKQLSVQYPRQPAI 960  
 DB 901 HTDSRYPVLSWVESRYNHYERTASDPQCSQNTILSSDRFGQPLKQLSVQYPRQPAI 960

QY 961 NLYPDTLPDKLLANSYDDQORQLRLTYQQSSWHLLTNNTVRVLGLPDSRSDIFTYGAEN 1020  
 DB 961 NLYPDTLPDKLLANSYDDQORQLRLTYQQSSWHLLTNNTVRVLGLPDSRSDIFTYGAEN 1020

QY 1021 VPAGGLNLELSDKNSLIADDKPREYLGQOKTAYTDGONTTPTQTPRQALIAFTETTVF 1080  
 DB 1021 VPAGGLNLELSDKNSLIADDKPREYLGQOKTAYTDGONTTPTQTPRQALIAFTETTVF 1080

QY 1081 NOSTLSAPNGSIPSDKLSLTTLEQAGYQOQTNVLPRTGEDKVVVAHGVTDYGTAAQFWRP 1140  
 DB 1081 NOSTLSAPNGSIPSDKLSLTTLEQAGYQOQTNVLPRTGEDKVVVAHGVTDYGTAAQFWRP 1140

QY 1141 QKQNTQLTGKITLIWDANCYVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNQHLITL 1200  
 DB 1141 QKQNTQLTGKITLIWDANCYVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNQHLITL 1200

QY 1201 DALGRPTILRFWGTENGKMTGYSSPEKASPPSDVNAALIELKKPLPVAQCVVAPESWM 1260  
 DB 1201 DALGRPTILRFWGTENGKMTGYSSPEKASPPSDVNAALIELKKPLPVAQCVVAPESWM 1260

QY 1261 PVLQKTFNRLAEOQWOKLYNARIITEDGRICTLAYRRWVQSOKAIIPOLISLLANGPRLP 1320  
 DB 1261 PVLQKTFNRLAEOQWOKLYNARIITEDGRICTLAYRRWVQSOKAIIPOLISLLANGPRLP 1320

QY 1321 PHSLLTTRDYDHDPEQIIRQQVVFSDGFRGLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380  
 DB 1321 PHSLLTTRDYDHDPEQIIRQQVVFSDGFRGLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380

QY 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDARSOKEAYADTHVYDPIGREIKV 1440  
 DB 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDARSOKEAYADTHVYDPIGREIKV 1440

QY 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAEVKKVKM 1474  
 DB 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAEVKKVKM 1474

RESULT 2  
 ADR21525  
 ID ADR21525 standard; protein; 1474 AA.  
 XX ADR21525;  
 AC AC  
 XX 04-NOV-2004 (first entry)  
 DT  
 DB Phototribus tcdB2 toxin SEQ ID NO:45.  
 XX toxin; insect; insecticidal; transgenic; pest control.  
 KW Phototribus luminescens.  
 OS  
 XX WO2004067727-A2.  
 FN  
 XX 12-AUG-2004.  
 PD  
 XX 07-JAN-2004; 2004WO-US0000394.  
 PF  
 XX 21-JAN-2003; 2003US-0441723P.  
 FR  
 XX (DOWC) DOW AGROSCIENCES LLC.  
 PA  
 XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
 XX WPI; 2004-580999/56.  
 DR  
 XX N-PSDB; ADR21524.  
 PT  
 XX Controlling or inhibiting an insect, useful for pest control, comprises  
 PT contacting the insect with effective amounts of a Protein A, a Protein B,  
 XX and a Protein C.  
 XX Claim 1; SEQ ID NO 45; 368pp; English.  
 XX

CC The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC protein A, a protein B, and a protein C. Each of the proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from XptA2wi,  
 CC XptA2wi, TcdA, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1b,  
 CC PpC1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wi, XptC1b,  
 CC (orf 6 long), PpC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a protein A, a protein B, and a protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC Photorhabdus luminescens TcdB2 toxin.

Sequence 1474 AA;

Query Match	100.0%;	Score 7901;	DB 8;	Length 1474;	
Best Local Similarity	100.0%;	Pred. No. 0;	Mismatches 0;	Gaps 0;	
Matches 1474;	Conservative	0;			
QY	1	MNSQDFISITELSLPKGGGAI	TGMEAL	TPGPGDMAALSPLPISAGRGYAPFTL	60
DB	1	MNSQDFISITELSLPKGGGAI	TGMEAL	TPGPGDMAALSPLPISAGRGYAPFTL	60
QY	61	SGAGNSPFGLGWDCNVMTIR	RRTHFGVPHYDE	DTDTFLGPEGEVLVADQPRDEST	120
DB	61	SGAGNSPFGLGWDCNVMTIR	RRTHFGVPHYDE	DTDTFLGPEGEVLVADQPRDEST	120
QY	121	LGATFTVTGYRSRLESF	SRLEYQPKTKT	DFWLIYSPDQVHLLGKSPQARISNPSQ	180
DB	121	LGATFTVTGYRSRLESF	SRLEYQPKTKT	DFWLIYSPDQVHLLGKSPQARISNPSQ	180
QY	181	TTQTAQWLLLEASVSSRGEQIYYQRAEDD	TGCEADEITHLQATAQRYLHI	VYVYGNRTAS	240
DB	181	TTQTAQWLLLEASVSSRGEQIYYQRAEDD	TGCEADEITHLQATAQRYLHI	VYVYGNRTAS	240
QY	241	ETLPGDGSAPQADWL	FYLPVDFYGBRSNNL	KTPPAFTTGSWLCRQDRFSRYEGFEIR	300
DB	241	ETLPGDGSAPQADWL	FYLPVDFYGBRSNNL	KTPPAFTTGSWLCRQDRFSRYEGFEIR	300
QY	301	TRRLCROVLYHHQLQALDSKITEHNGP	TLVSRLLIYNDSEAIASFLVFRVVGHEQGNV		360
DB	301	TRRLCROVLYHHQLQALDSKITEHNGP	TLVSRLLIYNDSEAIASFLVFRVVGHEQGNV		360
QY	361	VTLPPLELAYQDFSPRHHQPMQDVLANFNAIQRWQLVLDKGEGLPGLLYQDKGAWMYR			420
DB	361	VTLPPLELAYQDFSPRHHQPMQDVLANFNAIQRWQLVLDKGEGLPGLLYQDKGAWMYR			420
QY	421	SAORLGEIGSDAVTWEKQPI	SVLPSSQSNASLVNDINGDGLDQDWITGPGGLRGVHSQRPD		480
DB	421	SAORLGEIGSDAVTWEKQPI	SVLPSSQSNASLVNDINGDGLDQDWITGPGGLRGVHSQRPD		480
QY	481	GSWTRFTPLNALPVEYTHPRQA	LDLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ		540
DB	481	GSWTRFTPLNALPVEYTHPRQA	LDLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ		540
QY	541	SGDITLPPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGFQIPITLPGFSQ			600
DB	541	SGDITLPPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGFQIPITLPGFSQ			600
QY	601	PATFBNPAQVYLAIDL	DGSGPTDLIVVHTNRDLIFLNKSGNGFAEPVTLRPFEGRLRFDHTC		660
DB	601	PATFBNPAQVYLAIDL	DGSGPTDLIVVHTNRDLIFLNKSGNGFAEPVTLRPFEGRLRFDHTC		660
QY	661	QLOMADVQGLGVASLI	LSVPHMSPHHWC	CDLTNNKPKWLLNEMNNMNVHHTLRYRSSQF	720
DB	661	QLOMADVQGLGVASLI	LSVPHMSPHHWC	CDLTNNKPKWLLNEMNNMNVHHTLRYRSSQF	720

QY	721	WLDEKAAALTTGQTPVCYLPPPIIHTLMQTEDEBISGNKLVTTILRYARGAWDGRERERPG	780
DB	721	WLDEKAAALTTGQTPVCYLPPPIIHTLMQTEDEBISGNKLVTTILRYARGAWDGRERERPG	780
QY	781	FGYVEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEVWRDDQAFAGSPRPT	840
DB	781	FGYVEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEVWRDDQAFAGSPRPT	840
QY	841	TWQDNKDVPVLTPEDDNSRYWFNRALKGQLLRSELYGLDDSTNKHVPYTVTFRSQVRRLQ	900
DB	841	TWQDNKDVPVLTPEDDNSRYWFNRALKGQLLRSELYGLDDSTNKHVPYTVTFRSQVRRLQ	900
QY	901	HTDSRYPVLMSSVSVESRNYHYERIAADDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPAI	960
DB	901	HTDSRYPVLMSSVSVESRNYHYERIAADDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPAI	960
QY	961	NLYPDTLPDKLLANSYDDQOQRLRTYQOOSWHLTNNTVRLGLPDSSTRSDIFTYGAEN	1020
DB	961	NLYPDTLPDKLLANSYDDQOQRLRTYQOOSWHLTNNTVRLGLPDSSTRSDIFTYGAEN	1020
QY	1021	VPAGLNLLELLSDKNLSLIADDPKREYLGQOQKTAAYTDGQNTTPTLQTPTRQALIAFTETTVF	1080
DB	1021	VPAGLNLLELLSDKNLSLIADDPKREYLGQOQKTAAYTDGQNTTPTLQTPTRQALIAFTETTVF	1080
QY	1081	NQSTLSAFNGSIPSDKLSLTTLLEQAGYQOQNTYLPRTGEDKVVVAHHGYTDYGTAAQFWRP	1140
DB	1081	NQSTLSAFNGSIPSDKLSLTTLLEQAGYQOQNTYLPRTGEDKVVVAHHGYTDYGTAAQFWRP	1140
QY	1141	QKQSNLTQTKITLIWDANYCVVQTRDAAGLTTTSKYDWRFLTPVQLTDINDNOHLITL	1200
DB	1141	QKQSNLTQTKITLIWDANYCVVQTRDAAGLTTTSKYDWRFLTPVQLTDINDNOHLITL	1200
QY	1201	DALGRPITILRFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM	1260
DB	1201	DALGRPITILRFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPESWM	1260
QY	1261	PVLSQKTFNRLAEQDWKLYNARIITEDGRICCTLAYRRVWSQKAIPOILISLLNNGPRLP	1320
DB	1261	PVLSQKTFNRLAEQDWKLYNARIITEDGRICCTLAYRRVWSQKAIPOILISLLNNGPRLP	1320
QY	1321	PHSLTLTTRDYDHDPEQIRQOQVPSDGPGRLLQAAARHEAGMARQNRNEDGSLIINVQHT	1380
DB	1321	PHSLTLTTRDYDHDPEQIRQOQVPSDGPGRLLQAAARHEAGMARQNRNEDGSLIINVQHT	1380
QY	1381	ENRWAVTGRTEYDNKGQPIRTYQYFVFLNDWRVVSND SARQEKAYADTHVYDPIGREIKV	1440
DB	1381	ENRWAVTGRTEYDNKGQPIRTYQYFVFLNDWRVVSND SARQEKAYADTHVYDPIGREIKV	1440
QY	1441	ITAGWFRRLTFTPWFVTVNEDENDTAAAEVKVKVM	1474
DB	1441	ITAGWFRRLTFTPWFVTVNEDENDTAAAEVKVKVM	1474
RESULT 3			
ABM70222			
ID	ABM70222	standard; protein; 1476 AA.	
AC	ABM70222;		
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE		Photorhabdus luminescens protein sequence #3319.	
XX			
KW		Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;	
KW		detection; food; gene expression; plant; animal; microorganism; toxin;	
KW		antibiotic; biopesticide; virulence factor; disease model; plague;	
KW		whooping cough.	
OS		Photorhabdus luminescens.	
XX			
PN		WO200294867-A2.	
XX			

PD 28-NOV-2002.  
 XX 07-FEB-2002; 2002MO-IB003040.  
 XX 07-FEB-2001; 2001FR-00001659.  
 XX (INSP ) INST PASTEUR.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 XX Buchrieser C;  
 XX WPT; 2003-148459/14.  
 XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX Claim 2; SEQ ID NO 3319; 1205pp; French.  
 XX The invention relates to the isolation of genes and their encoded  
 XX proteins from Photorhabdus luminescens. The isolated sequences are  
 XX sources of probes and primers for detecting the genome of P. luminescens  
 XX and related species; to study polymorphisms; for gene analysis and for  
 XX detection/amplification of the genes. Antibodies (Ab) raised against the  
 XX polypeptides encoded by the genes are used for detection/identification  
 XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 XX carry a gene-containing vector are used to select compounds that  
 XX modulate, regulate, induce or inhibit expression of the genes in plants,  
 XX animals or microorganisms other than P. luminescens and are able to alter  
 XX response or sensitivity to toxins and antibiotics produced by P.  
 XX luminescens. Cells transformed to express the genes are useful for  
 XX recombinant production of the proteins, particularly toxins and  
 XX antibacterials useful as insecticides, bactericides and fungicides. The  
 XX genes, proteins, vectors containing the genes and Ab are also useful  
 XX therapeutically (to treat microbial infection by bacteria or fungi that  
 XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 XX biopesticides. Other uses of the genes and the proteins are as virulence  
 XX factors and for identifying targets of human diseases for which P.  
 XX luminescens is a model (particularly plague and whooping cough). This  
 XX sequence represents one of the isolated P. luminescens proteins  
 XX  
 SQ Sequence 1476 AA;  
 Query Match 92.2%; Score 7285.5; DB 6; Length 1476;  
 Best Local Similarity 91.7%; Pred. No. 0;  
 Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;  
 1 MONSDPFTITSLPKGGGCAITGMEALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60  
 1 MONSDPFTITSLPKGGGCAITGMEALTPGPDGMAALSPLPISAGRGYAPALSALNYN 60  
 61 SGAGNSPFLGWCNVMTIRRTTHFGVPHYDETDFLGPGEVLVADQPRDESTLQGIN 120  
 61 SGAGNTPFLGWCNVMTIRRTTHFGVPHYNETDFLGPGEVLVADQPRDESTLQGIN 120  
 121 LGATFTVTGYSRLESFSLLEYQPKTKGDFWLIYSPDGQVHLLGKSPQARISNPSQ 180  
 121 LGATFTVTGYSRLESFSLLEYQPKTKGDFWLIYSPDGQVHLLGKSPQARISNPSQ 180  
 181 TTQTAQWLLASVSRSRQIYYQRAEDDTGCEADEITHLQATAQRYLHIVYIGNRTAS 240  
 181 TTQTAQWLLASVPHGQIYYQRAEDNLCETNEITHLQATAQRYLHIVYIGNRTAS 240  
 241 ETLPGLDGASPSQADWLFLYVDFYGERNNLTKTPPAFTTGSWLCRQDRFSRYEYGFIR 300  
 241 ETLPLNGRAPSQADWLFLYVDFYGERNNLTKMPPAFTATGNWLCRQDRFSRYEYGFALR 300  
 301 TRRLCROVLMYHLLQALDSKITEHNGPTLVSRLLIINYDESAIATFLVPRVRGHEODGNV 360  
 301 TRRLCROQLMYHLCQALDNKIKENHGPTLVSRLLIINYDESAIATFLVPRVRGHEODGTA 360  
 361 VTLPPLELAYQDFSPRHARHPQMDVLANFNAIQRWQLVDLKGEGPLGLLYQDKGAMWYR 420

Db VTLPPLELAYQDFSPQHNTHQPMVDLANFNAIQRWQLVDLKGEGPLGLLYQDKNAWMYR 420  
 Qy SAQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVDINGDQGLDQWVITGPGRGVHSORPD 480  
 Db SAQRLGEIGSDVTVWEKIQPLSVIPSLQSNASLVDINGDQGLDQWVITGPGRGVHSORPD 480  
 Qy GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLDVLIGPKSVRLYANTRDGFAGKDVVQ 540  
 Db GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLDVLIGPNSVRLYANTRDGFAGKDVVQ 540  
 Qy SGGITLVPVPGADPRKLVAFSDVLGSGQAHLEVESATKVTWCPNIGRGRFGQPITLPGFSQ 600  
 Db SGGITLVPVPGADPRKLVAFSDVLGSGQAHLEVESATKVTWCPNIGRGRFGQPITLPGFSQ 600  
 Qy PATFENPAQVYIADLDGSGPTDLIYVHTNRDLIDFLNKGSGNGFAEPVILRPFEGILRFDHTC 660  
 Db PEAFENPAQVYIADLDGSGPTDLIYVHTNRDLIDFLNKGSGNGFAEPVILRPFEGILRFDHTC 660  
 Qy QLOWADVOGLGVASLILSVPHMSPHWRCDLTNNKPWLLNEMNNMGMVHHTLYRVSQQF 720  
 Db QLOWADVOGLGVASLILSVPHMTPHWRCDLTNNKPWLLNEMNNMGMVHHTLYRVSQQF 720  
 Qy WLDEKAAALTTGQTPVCYLPPPIHTLMQTEDEISGNKLVTTILRYARGAWDGREREPRG 780  
 Db WLDEKAAALVAGQTPVCYLPPPIHTLMQTEDEISGNKLVTTILRYARGAWDGREREPRG 780  
 Qy FGYYEQTDSHOLAQGNAPERTPPALTKNWTATGLPVIDNALSTEVYR-DDOAEAGFSRPF 839  
 Db FGYYEQTDSHOLAQGNASERTPPAMTKNWTATGLPVIDNTLSTEVYRWDGNQAFAGFSRPF 840  
 Qy TTWQDNKDVPLTPEDDNRVYFNRALQGLRLSBLYLGLDDSTNKHVPYTVTFEPRSQVRL 899  
 Db TTWQDNKDVPLTPEDDNRVYFNRALQGLRLSBLYLGLDDSTNKHVPYTVTFEPRSQVRL 900  
 Qy QHTDSRYPLVWSSVVEGRNHYERIASDPQCSQMITLSSDRFGPGLKSLVQYPRRQOPA 959  
 Db QHTDSRYPLVWSSVVEGRNHYERIASDPQCSQMITLSSDRFGPGLKSLVQYPRRQOPA 960  
 Qy INLYPDTLPDKLLANSYDDQORQLRLTYQSSWHHLTNTVTVLGLPSTSDIFTYGAE 1019  
 Db ISPYDPTLPDELLANSYDEQORQLRLTYQSSWHHLTNTVTVLGLPSTSDIFTYGAE 1020  
 Qy NVPAAGLNLELLSDKNSLIADDDKPREVLGQOKTAYTDGQNTTPTLQTPRQALIAFTETTV 1079  
 Db NVPAAGLNLELLSDKNSLIADDDKPREVLGQOKTAYTDGQNTTPTLQTPRQALIAFTETTV 1080  
 Qy FNOSTLSAFNGSIPSDKSLSTTLEAGYQOQNTYLPFRGTGDKVVAHGYTDYGTAAQFWR 1139  
 Db FNOSTLSAFNGSIPSAQLSTTLEAGYQOQNTYLPFRGTGDKIWAHGYTDYGTAAQFWR 1140  
 Qy POKOSNTOLTGKITLINDANYCVVVTQTRDAAGLTTSYAKYDWRFLTPVOLTINDNQHILIT 1199  
 Db POKOSNTOLTGKITLINDANYCVVVTQTRDAAGLTTSYAKYDWRFLTPVOLTINDNQHILIM 1200  
 Qy LDALGRPITLAFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCVYAPESW 1259  
 Db LDALGRPITLAFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCVYAPESW 1260  
 Qy MPVLSQKTFNRLAODWQKLNARIITEDGRICTIAYRRWVQSOKAIQPLISLLNNGRL 1319  
 Db MPVLSQKALNRLAODWQKLNARIITEDGRICTIAYRRWVQSOKATPOLTSLNNGPHL 1320  
 Qy PPHSLTITTDYRDHDPQOIRQOVVFSDFGRLLQAAARHAGMARQORNEGSLIINVQH 1379  
 Db PPHSLTITTDYRDHDPQOIRQOVVFSDFGRLLQAAARHAGMARQORNEGSLIINVQH 1380  
 Qy TENERWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVDPITGREIK 1439  
 Db TENERWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVDPITGREIK 1440  
 Qy VITAKGFRRTLTPFWFTVNEDENDTAAEVKVKVM 1474  
 Db VITAKGFRRTLTPFWFTVNEDENDTAAEVKVKVM 1475

RESULT 4

ID ABM70230 standard; protein; 1477 AA.

XX AC ABM70230;

XX DT 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #3327.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis; detection; food; gene expression; plant; animal; microorganism; toxin; antibiotic; biopesticide; virulence factor; disease model; plague; whooping cough.

XX OS Photorhabdus luminescens.

XX PN WO200294867-A2.

XX PD 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX PR 07-FEB-2001; 2001FR-00001659.

XX PA (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A; Buchrieser C;

XX PT WPI; 2003-148459/14.

XX PS Genomic sequence of Photorhabdus luminescens and encoded polypeptides, useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX SQ Claim 2; SEQ ID NO 3327; 1205pp; French.

CC The invention relates to the isolation of genes and their encoded proteins from Photorhabdus luminescens. The isolated sequences are sources of probes and primers for detecting the genome of P. luminescens and related species; to study polymorphisms; for gene analysis and for detection/amplification of the genes. Antibodies (Ab) raised against the polypeptides encoded by the genes are used for detection/identification of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that carry a gene-containing vector are used to select compounds that modulate, regulate, induce or inhibit expression of the genes in plants, animals or microorganisms other than P. luminescens and are able to alter response or sensitivity to toxins and antibiotics produced by P. luminescens. Cells transformed to express the genes are useful for recombinant production of the proteins, particularly toxins and antibacterials useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and Ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to P. luminescens-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which P. luminescens is a model (particularly plague and whooping cough). This sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 1477 AA;

Query Match 76.8%; Score 6065.5; DB 6; Length 1477;

Best Local Similarity 76.3%; Pred. No. 0;

Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

QY 1 MNSQDFISITSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGVAPFTLNYN 60

DB 1 MNSQTFVAELSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGVSPSLTNS 60

QY 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

Db	61	SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVAALNENQADIRSES	120
QY	115	TLOGINLGATFTVTGYSRLESFHSRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR	174
Db	121	SLQGINLGEFTVTGYSRLESFHSRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR	180
QY	175	ISNPSQTTQTAQWILLEASVSRSRGQIYYQYRAEDDTGCEADEITHLQATAQRYLHYVY	234
Db	181	ISNPLNVSQTAQWILLEASVSRSRGQIYYQYRAEDDTGCEADEITHLQATAQRYLHYVY	240
QY	235	GNRTASETLPLGLDGSAPSOADWLVFVDFYGERSNLKTTPAFSTTGSWLCRQDRFSRYE	294
Db	241	GNLTASEVFTPLNGDDPLKSGWLFCLVPDYGERKNSLSEIPFKASSLWLCRQDRFSRYE	300
QY	295	YGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAITASTLVFVRRVGH	354
Db	301	YGFELTRRLCRQILMFHRLQTLTSGQAKGDDDEPALVSRLLIDYDENAVISTLVSVRRIGH	360
QY	355	EQDGNVVTLPPELAYQDFSPRHHAHQPMQMDVLANFNAIQRWQLVDLKGELPGLLYQDK	414
Db	361	EDNNTVLSLPPLELAYQDFPEQKARWQSMQMDVLANFNAIQRWQLVDLKGELPGLLYQDK	420
QY	415	GAWYRSQRLGEIGSDAVTWKMQPLSVIFSLQSNASLVDFINGDQGLDWITGFLRGY	474
Db	421	NGWYRSQRLGEIGSDAVTWKMQPLSVIFSLQSNASLVDFINGDQGLDWITGFLRGY	480
QY	475	HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTDRGFAK	534
Db	481	HSQRPDGSWTRFTPLDALPIEYSHPRAQLADLMGAGLSDLVLIGPKSVRLYANNRDRGFTQ	540
QY	535	GKDVQSGDITLPLVPGADPRKLVAFSDVLGSGQAHLEVEVSATKVTCPNPLGRGRGQIT	594
Db	541	GRDVQSGDITLPLVPGADPRKLVAFSDVLGSGQAHLEVEVSATKVTCPNPLGRGRGQIT	600
QY	595	LPGFSQATENPRAQVLAIDLGGSPDTLIIVHTNRDLDFLNKSGNGFAEPVTLRFPBGL	654
Db	601	LPGFSQADNPNDRVHLADLGGSPADLIIVHTDRLEIFKNESGNSFAKFTLRFPPGL	660
QY	655	RFDTTCQLQWADVQGLGVALSLVSPHMSPHHRCDLTNMKPWLNLNENNMNVHHTLRY	714
Db	661	RFDDTCQLQWADVQGLGVALSLVSPHMAPHHRCDLTNAPFWLLSEMMNMGAAHTLHY	720
QY	715	RSSQFMDLDEKAAALATGCTQTPVCVLPPIHTLQWTEDEISGNKLVTLRYARGAWGR	774
Db	721	RSSQFMDLDDKAAALATGCTQTPVCVLPPEVHTLQWTEDEISGNKLVTLRYAHGAWGR	780
QY	775	EREPRGFGYVQTDHSHOLAQGNAPERTPPALTQWYATGLPVIDNALSTEYWR-DDQAFA	833
Db	781	EREPRGFGYVQTDHSHOLAQGNAPERTPPALTQWYATGLPVIDNALSTEYWR-DDQAFA	840
QY	834	GFSPRFTTWQDNKDVPLTPEDDNRYSRYWFRNALKGQLLSELYGLDDSTNKHVPYTVTEFR	893
Db	841	GFTPRFTRWKSGKDVLPATPENDNLYWFRNALKGQLLSELYGLDDSTNKHVPYTVTEFR	900
QY	894	SQVRLQHTDSRYVPLVSSVVSRYNHYERTASPPQCSQNTLSDDRGPQLKQLSVQYP	953
Db	901	PQVRLQDGTSTASVPLVSSVVSRYNHYERTASPPQCSQNTLSDDRGPQLKQLSVQYP	960
QY	954	RQOPAINLYPDTLPDKLANSYDDQQLALTYQSSSMHHLTNNTVRLVGLPDSTRDI	1013
Db	961	RRNFTNTPYPTLPTLTFASSYDDQQLALTYQSSSMHHLTNNTVRLVGLPDSTRDI	1020
QY	1014	FTYGAENVPAAGLNLLELSDKNSLIADDKPREYLGQOQKTYATDQNTTFLQTPTRQALIA	1073
Db	1021	FTYDAKQVVDGLNLEALCAENSLIADDKPREYLNQOQKTYATDQNTTFLQTPTRQALIA	1080
QY	1074	FTETTVFNQSTLSAPNGSIPSDKLSSTLLEQAGYQOQNTLFPRTGDKVVAHGHVDTYGT	1133
Db	1081	FTETAVLTESLSLAFDGGITPDELPGILTQAGYQOQEPYLPFRPTGDKVVAHGHVDTYGT	1140
QY	1134	AAQWRFPQKQNTQLTGKITLIWIDANYCVVQTRDAAGLTTSKAYDMRFLTPVQLTDLND	1193



Db 1141 EAQFWRPVAQRNTLLTGKTKTLQWDTHYCVITQTQDAAGLTVLVANDWRFLTPVQLTDIND 1200  
 Qy 1194 NQHLLITDALGRPTTLRFWGTENGWYGSPEKASFPSPSDVNAATLKKPLPVACOV 1253  
 Db 1201 NVHLITLIDALGRPVTRQFENGWYGSPEKFPSPDIDINTALTALGPLVACQCLV 1260  
 Qy 1254 YAPESWMPVLQKTFNRLAEQDWOKLVNARIITBDGRICLTAYRRWVQSKAIPQLISLL 1313  
 Db 1261 YAPSWMPVLSQEFNTLTQEQTLRLDLRIITEDWRICALARRWLSQASFLVNL 1320  
 Qy 1314 NNGPRLPHSLTLLTDRYDHPDEQIRQOVVFSDFGRLLQAAARHAGMARQNEGSL 1373  
 Db 1321 TNSIGLPHNLTLTPDRYDRDSGQIIHQVAFSDGFRLLQASVRHEAGAWQRNQDGL 1380  
 Qy 1374 IINVOHTENRVAVTGRTEYDNKGPIRTYQPYFLNDWRYUSNDSAROEKAYADTHYDP 1433  
 Db 1381 VTKMEDTKRVAVTGRTEYDNKGPIRTYQPYFLNDWRYUSNDSAR--KGAYADTHYDP 1438  
 Qy 1434 IGRBIKVTAKGWFRRTLTFTWFTVNEDENDTAAEV 1469  
 Db 1439 IGRIRVITAKGWLRSQSFVFWFVSEDENDTAAEV 1474

## RESULT 5

AA333730  
 ID AA333730 standard; protein; 1481 AA.

AC AA333730;

XX 09-NOV-1999 (first entry)

XX Photorhabdus luminescens 1481 amino acid insecticidal toxin.

XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX Photorhabdus luminescens.

XX WO9942589-A2.

XX 26-AUG-1999.

XX 18-FEB-1999; 99WO-EP001015.

XX 20-FEB-1998; 98US-00027080.

XX 20-JAN-1999; 99US-0116439P.

XX (NOVS ) NOVARTIS AG.

XX (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MW;  
 PI Chen JS;

XX WPI; 1999-527479/44.

XX N-PSDB; AA206831.

XX New nucleic acid from Photorhabdus luminescens encoding insecticidal  
 PT toxins, used for making resistant transgenic plants.

XX Claim 26; Page 140-145; 148pp; English.

XX This sequence represents a 1481 amino acid insecticidal toxin from  
 CC Photorhabdus luminescens. It is one of three insecticidal toxins  
 CC (AA333728-Y33730) encoded by open reading frames (orfs) in a 38kb  
 CC fragment of P. luminescens DNA (AA206831). This sequence is encoded by  
 CC orf2. P. luminescens is a member of the Enterobacteriaceae family and is  
 CC a symbiotic bacterium of nematodes of the genus Heterorhabditis. The  
 CC nematodes colonise insect larvae, kill them, and their offspring feed on  
 CC the dead larvae. However, the insecticidal agents are produced by P.  
 CC luminescens rather than the nematodes. The toxins have activity against  
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni), European  
 CC Corn Borer (Ostrinia nubilalis) and Fall Armyworm (Spodoptera perfrida)  
 CC and also against Coleopteran insects (e.g., Colorado Potato Beetle,  
 CC Leptinotarsa decimlineata). In addition the toxins are active against

CC strains resistant to known insecticides. The DNA sequence can be used to  
 CC generate transgenic plants of various species that are resistant to  
 CC economically important insect pests and also for recombinant production  
 CC of the toxins for use as insecticides

XX Sequence 1481 AA;

Query Match 76.5%; Score 6042; DB 2; Length 1481;  
 Best Local Similarity 76.1%; Pred. No. 0;  
 Matches 112; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy 1 MONSQDSITELSLPKGGGALTGMGEALTTPGPDGMAALSPLPISAGRGYAPFTLNYN 60

Db 1 MONSQDSITELSLPKGGGALTGMGEALTTPGPDGMAALSPLPISAGRGYAPFTLNYN 60

Qy 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDEDTDFLGPGEVLVVA-----DQPRDES 114

Db 61 SGTGNSPFLGWCNVMVIRRRTHFGVPHYDEDTDFLGPGEVLVVAALNEAGADIRSES 120

Qy 115 TLQGINLGATFTVTGYSRLESHPFSLRYWQPKTKTDFWLIYSPDGQVHLLCKSPQAR 174

Db 121 SLOGINLGMFTVTGYSRLESHPFSLRYWQPKTKTDFWLIYSPDGQVHLLCKSPQAR 180

Qy 175 ISNPSQTTQTAQMLLEASVSRSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 234

Db 181 ISNPLNVNQTAAQMLLEASVSRSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 240

Qy 235 GNTASSETLPGDGSAPSQADWLPYLVDFYDGERSNLUKTPPAFSTTGSWLCRQRFERYE 294

Db 241 GNLTASEVFPFTLNGDDPLKSGWLFCLVFDYGERKNSLSEMPFPKATSNWLCRQRFERYE 300

Qy 295 YGFEIRTRRLCRQVLMYHHLQALDSKITTEHNGPTLVSRLLIINYDESAIASTLVPRVVG 354

Db 301 YGFAIRTRRLCRQVLMYHHLQALDSKITTEHNGPTLVSRLLIINYDESAIASTLVPRVVG 360

Qy 355 EQDG-NVVTLPPLPPLAYQDFSPRHHAHQPMVLANFNAIQRWOLVDLKGEGPLGLYQD 413

Db 361 EQDGTVALPPLPPLAYQDFSPRHHAHQPMVLANFNAIQRWOLVDLKGEGPLGLYQD 420

Qy 414 KGAWYRSARLGEIGSDAVTWEKMQPLSVIPSLQSNASLVINDGDLQDLVITGPGIRG 473

Db 421 KNGWYRSARLGEIGSDAVTWEKMQPLSVIPSLQSNASLVINDGDLQDLVITGPGIRG 480

Qy 474 YHSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGA 533

Db 481 YHSQHPDGSWTRFTPLHALPIETHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGA 540

Qy 534 KGKDVVQSGDITLFPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLGRGRFGQPI 593

Db 541 EGRDVVQSGDITLFPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLGRGRFGQPI 600

Qy 594 TLPFGSQPATEFNPQAVYLADLQSGGPTDLIYVHTNRLDIPLNKSNGGFAEPVTLRPEG 653

Db 601 TLPFGSQPATEFNPQAVYLADLQSGGPTDLIYVHTNRLDIPLNKSNGGFAEPVTLRPEG 660

Qy 654 LRFDHTCOLQWADVQGLGVASLLISVPHMSPHWRCDLTNKPWLLMNMNMVGHVTLR 713

Db 661 LRFDHTCOLQWADVQGLGVASLLISVPHMSPHWRCDLTNKPWLLMNMNMVGHVTLR 720

Qy 714 YRSSSQFWLDEKAAALTGTQTPVCYLPFPIHTLMQTEDEISGNKLVTLRLYARGAWDG 773

Db 721 YRSSSQFWLDEKAAALTGTQTPVCYLPFPIHTLMQTEDEISGNKLVTLRLYARGAWDG 780

Qy 774 RERFRFGVGVQEDSHQLAQGNAPERTPPALTKQWTATGLPVIDNALSTEYWR-DOAF 832

Db 781 RERFRFGVGVQEDSHQLAQGNAPERTPPALTKQWTATGLPVIDNALSTEYWR-DOAF 840

Qy 833 AGSPRFTTQDNKQVPLTPEDNSRYFNRLKGQLLRSELVGLDSDTNKHPVTVTFEF 892

Db 841 TGFTPHFLWKRGKDVPLTPEDDNLNWLNRALKGQPLRSELVGLDSDTNKHPVTVTFEF 900

Qy 893 RSQVRLQHTDSRYFVLWSSVSRNHYERIASDQCSQNTLSSRDFGQPLKQLSVQY 952



Db 901 RPQVRLQDNVTLSPVLNWSVRSYHYRIISDPQCNQDITLSSDIFGQPLKQVSQY 960  
Qy 953 PRQOPAINLYPDTLPDKLANSYDDQQRQLTYQQSSWHHLTNTNTRVLGLPDRSD 1012  
Db 961 PRENKPTNPYPDTLPDTLPASSYDDQQLRLTYQQSSWHHLTANELRVLGLPDRSD 1020  
Qy 1013 IFTYGAENVVAGLNLBLLSKNSLIADKPREYLGQKQYAY----TQGNVTPLOTPR 1068  
Db 1021 AFTYDAKHVPVGLNLEALCAENSIIADKPREYLNQOQRTFTYDQKTDGKNPTPLKTPR 1080  
Qy 1069 QALIAFTETTVNQSTLAFNGSIPSDKLTSTLEAGYQOQNTYLPFRPTGEDKVVVAHGY 1128  
Db 1081 QALIAFTETTVNQSTLAFNGSIPSDKLTSTLEAGYQOQNTYLPFRPTGEDKVVVAHGY 1140  
Qy 1129 TDYGTAAQFWRPQKQNSQTLQTKITLIWDANYCVVQTRDAAGLTTSKRYDMRFLTPVQL 1188  
Db 1141 TDYGTGVQFWRPQKQNSQTLQTKITLIWDANYCVVQTRDAAGLTTSKRYDMRFLTPVQL 1200  
Qy 1189 TDINDNHLITLDALGRITLRFWGTENGKWTGYSPEKASFPSPDVNAALIKKPLPV 1248  
Db 1201 TDINDNHLITLDALGRITLRFWGTENGKWTGYSPEKASFPSPDVNAALIKKPLPV 1260  
Qy 1249 AOCVYAPESWNPVLSQKTFNRLABODMQLYNARIITEDGRICTLAVRRVQSKAIPQ 1308  
Db 1261 AOCVYAPESWNPVLSQKTFNRLABODMQLYNARIITEDGRICTLAVRRVQSKAIPQ 1320  
Qy 1309 LISLNNPRLPPLSHLITTDYDHPDQIQVQVVFSDGFRLLIQAARHEAGMARQNR 1368  
Db 1321 LVKLLTNSIGLPPHNLMLATDRYDRDSEQQIRQVAFSDGFRLLIQAARHEAGMARQNR 1380  
Qy 1369 EDGSLINQVHTENRWATGTEYDNKQOPIRTYQYFINDWRVYNSDASQEKAYADT 1428  
Db 1381 QDGLSVTQWEDTKRWAITGTEYDNKQOPIRTYQYFINDWRVYNSDASQEKAYADT 1438  
Qy 1429 HYDPIGRIEIKVITAKGFRITLTPFTVNEDENDTAAE 1468  
Db 1439 HYDPIGRIEIKVITAKGFRITLTPFTVNEDENDTAAE 1478

RESULT 6  
ABG32652  
ID ABG32652 standard; protein; 1476 AA.  
XX AC  
XX AC  
XX DT  
XX DT  
XX DE  
XX DE  
XX KW  
XX KW  
XX OS  
XX OS  
XX FN  
XX FN  
XX PD  
XX PD  
XX PF  
XX PF  
XX PR  
XX PR  
XX PA  
XX PA  
XX PA  
XX PA  
XX PI  
XX DR  
XX DR  
XX DR  
XX PT

Gene; ds; toxin A; toxin B; TcdA; protoxin; TcdB; TccC2; transgenic; monocot cell; dicot cell; oral toxin; insect; pest; TcBa.  
Photorhabdus luminescens.  
US2002078478-A1.  
20-JUN-2002.  
26-MAR-2001; 2001US-00817514.  
24-MAR-2000; 2000US-0191806P.  
(FFRE/) FRENCH-CONSTANT R H.  
(BOWE/) BOWEN D.  
(ROCH/) ROCHELEAU T A.  
(WATE/) WATERFIELD N R.  
Ffrench-Constant RH, Bowen D, Rocheleau TA, Waterfield NR;  
WPI; 2002-655379/70.  
N-PSDB; ABS52584.  
Novel nucleic acid sequences which encode genes, tcdB and tccC2 from

Photorhabdus luminescens W-14, useful in heterologous expression of orally active insect toxins.  
Claim 1; Page 20-23; 40pp; English.  
The invention discloses an isolated nucleic acid that encodes TcdB or TccC2 from Photorhabdus luminescens W-14. Also disclosed is a transgenic monocot or dicot cell and a transgenic plant (including the seeds) both with genomes comprising tcdB and tccC2 nucleic acids. The nucleic acids are useful for producing Toxin A or B of P. luminescens W-14 in a heterologous host and for encoding TcdB or TccC2 for producing an orally active insect toxin in a host, where the host also expresses TcdA or TcBa from P. luminescens W-14. Heterologous expression of Toxin A does not afford the level of oral toxicity to insects as that of the native toxin, but the coexpression increases this toxicity. The transgenic plants expressing effective amounts of the toxins protect themselves from insect pests. When the insects feeds on the transgenic plant it also ingests the toxins and this deters the insect from further biting into the plant and may even harm or kill the insect. The sequence presented is the P. luminescens (W-14) TcdB protein  
Sequence 1476 AA;  
Query Match 76.3%; Score 6031.5; DB 5; Length 1476;  
Best Local Similarity 75.9%; Pred. No. 0;  
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;  
Qy 1 MONSQDSITELSLPKGGGALTGMEALTTPGPGMAALSPLPISAGRGYAPFTLNN 60  
Db 1 MONSQDSITELSLPKGGGALTGMEALTTPGPGMAALSPLPISAGRGYAPFTLNN 60  
Qy 61 SGAGNSPFLGWDGNVMTIRRRTHFGVPHYDEDTFLGPEGEVLVVA-----DQPRDES 114  
Db 61 SGTGNSPFLGWDGNVMTIRRRTHFGVPHYDEDTFLGPEGEVLVVAALNAQADIRSES 120  
Qy 115 TLOGINLGATFTVTCYRSRLESHPESRLEYQPKTKTDFWLIYSPGOVHLGKSPQAR 174  
Db 121 SLOGINLGATFTVTCYRSRLESHPESRLEYQPKTKTDFWLIYSPGOVHLGKSPQAR 180  
Qy 175 ISNPSQTTQTAQMLLEASVSRSRQIYYQVRAEDDTGCEADEITHHQAQRYLHIVY 234  
Db 181 ISNPLNVNQAQMLLEASISHSRQIYYQVRAEDAGCETDELAHSPATVQRYLQTVHY 240  
Qy 235 GNRITASLPLGLDGSAPSQADWLFPYVDYGERSNKLKTPPAFTTGSWLCRQDRFSRYE 294  
Db 241 GNLTASDVFPPLANGDDPLKSGMFCVFDYGERKNSLSEMPLEKATGNLWLCRDRFSRYE 300  
Qy 295 YGPEIRTRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESATIASTLVFVRVGH 354  
Db 301 YGFELRTRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESATIASTLVFVRVGH 360  
Qy 355 EQDGNVVTLPPELAYQDFSPRHHAHQPMQMDVLANFNAIQRWQLVLDKGEGLPGLLYQDK 414  
Db 361 EDNNTVTALPPELAYQDFSPRHHAHQPMQMDVLANFNAIQRWQLVLDKGEGLPGLLYQDK 420  
Qy 415 GAWYRSQRLGEIGSDAVTWKQPLSVIPSLQSNASLVNDINGDGLDQWITGPGLRGY 474  
Db 421 NGWYRSQRLGEIGSDAVTWKQPLSVIPSLQSNASLVNDINGDGLDQWITGPGLRGY 480  
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVIGPKSVRLVANTRDGPAP 534  
Db 481 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVIGPKSVRLVANTRDGPAP 540  
Qy 535 GKDVVQSGDITLPPVPGADPRKLVAFSVGLSGQAHLEVVSATKVTCTWPNLGRGFGQPI 594  
Db 541 GRDVVQSGDITLPPVPGADPRKLVAFSVGLSGQAHLEVVSATKVTCTWPNLGRGFGQPI 600  
Qy 595 LPGFSQAPTEFNPQAVYLAADLDSGPTDLIYVHTNRDLDFUNKSGNGFAEPVTRFPEGL 654  
Db 601 LPGFSQAPTEFNPQAVYLAADLDSGPTDLIYVHTNRDLDFUNKSGNGFAEPVTRFPEGL 660  
Qy 655 RFDHTCOLQADVQGLGVASILLSPHMSPHHWFCDLTNNKDWLNEMNNMNVHHTLRY 714

Db 661 RFDDTCQLQVADVOGLGVSVLLSVPHMA PHHWRCDLTNAK PWWLLSETNNMNGANHTLIH 720  
 Qy 715 RSSQFMDLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTTLYRARGANDGR 774  
 Db 721 RSSVQFMDLDEKAAALTCQTPVCYLPPPVHTLMOTETEDISGNKLVTTLYRARGANDGR 780  
 Qy 775 EREPRGFGYVQTDHSHOLAQGNAPERTPPALTKKNYATGLPVIDNALSTEVWR-DDQAFPA 833  
 Db 781 EREPRGFGYVQTDHSHOLAQGNAPERTPPALTKKNYATGLPVIDNALSTEVWR-DDQAFPA 840  
 Qy 834 GFSPRFTWQNDKDVPLTPEDDNRKRYFNRAKQQLRSSELYGLDDSTNKHVPYVTVTEFR 893  
 Db 841 GFTFRFTLWKEGKDVPLTPEDDNLNLRALKQPLSELYGLDGSAAQOIPYTVTESR 900  
 Qy 894 SQVRLQHTDSRYVPLWSSVVESSRYHYERIASDPQCSQNTLSSDRFGQPLKLSVQYYP 953  
 Db 901 PQVRLQDQATVSPVWASVVESSRYHYERIASDPQCSQNTLSSDRFGQPLKLSVQYYP 960  
 Qy 954 RROQPAINLYEDTLDPDKLLANSYDDQORQLRLTYQSSWHHLTNTVTVLGLPDRSTRDI 1013  
 Db 961 RNNKPTNPNYEDTLDPDKLLANSYDDQORQLRLTYQSSWHHLTNTVTVLGLPDRSTRDI 1020  
 Qy 1014 FTYGAENVPAAGLNLLELSDKNSLIADDPREYLGQKQATYTDGQNTTLPLOTPTROALIA 1073  
 Db 1021 FTYDAKQVPVDGLNLETLCAENSLIADDPREYLNQORTFTYTDGKNQTPKLTPTROALIA 1080  
 Qy 1074 FTETTVNQSTLSPFNGISPSDKLSTLLEQAGYQQTNYLPPTGDEKQWVAHHGYTDYGT 1133  
 Db 1081 FTETAVLTESLSPFNGISPSDKLSTLLEQAGYQQTNYLPPTGDEKQWVAHHGYTDYGT 1140  
 Qy 1134 AAQFWRPQKQNTOLTGKTLIWDANVCVVQTRDAAGLTTSYAKYDWRFLTPVOLTIND 1193  
 Db 1141 EAQFWRPQKQNTOLTGKTLIWDANVCVVQTRDAAGLTTSYAKYDWRFLTPVOLTIND 1200  
 Qy 1194 NOHLITLDALGRPTLRFWGTENGKMTGYSPKASFPSPDNNAAIELKKPLVQAQCV 1253  
 Db 1201 NVHLITLDALGRPTLRFWGTENGKMTGYSPKASFPSPDNNAAIELKKPLVQAQCV 1260  
 Qy 1254 YAPESWNPVLSQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWQSKAIQPLISLL 1313  
 Db 1261 YAPESWNPVLSQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWQSKAIQPLISLL 1320  
 Qy 1314 NNGPRLPHSLTLTDRYDHPDQOIRQOVVPSDFGRLQLQAARHEAGMARQNRDGS 1373  
 Db 1321 TNSGLPPLHSLTLTDRYDHPDQOIRQOVVPSDFGRLQLQAARHEAGMARQNRDGS 1380  
 Qy 1374 IINYQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSARQEKAYADTHYDP 1433  
 Db 1381 VTKVENTKTRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVSNDSARQEKAYADTHYDP 1438  
 Qy 1434 IGRBKIVITAKGWFRTLTFTWFTVNEDENDTAAE 1468  
 Db 1439 IGRBKIVITAKGWFRTLTFTWFTVNEDENDTAAE 1473

## RESULT 7

ADL21576  
 ID ADR21576 standard; protein; 1476 AA.

XX AC ADR21576;  
 XX DT 04-NOV-2004 (first entry)  
 XX DE Photorhabdus TcdB1 toxin.  
 XX KW toxin; insect; insecticidal; transgenic; pest control.  
 XX OS Photorhabdus luminescens.  
 XX FN MO2004067727-A2.  
 XX PD 12-AUG-2004.  
 Qy 415 GAWWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQGLDWVITGFLRGY 474

PF 07-JAN-2004; 2004WO-US0000394.  
 XX 21-JAN-2003; 2003US-0441723P.  
 XX (DOWC) DOW AGROSCIENCES LLC.  
 XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
 XX WPI; 2004-580999/56.  
 DR N-PSDB; ADR21502.  
 XX Controlling or inhibiting an insect, useful for pest control, comprises  
 PT contacting the insect with effective amounts of a Protein A, a Protein B,  
 XX and a Protein C.  
 PS Claim 1; SEQ ID NO 22; 368pp; English.  
 XX The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from XptA1wi,  
 CC XptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1xb,  
 CC PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wi, XptC1xb, PptC1  
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC Photorhabdus luminescens TcdB1 toxin.

XX SQ Sequence 1476 AA;

Query Match 76.3%; Score 6025.5; DB 8; Length 1476;  
 Best Local Similarity 75.9%; Pred. No. 0;  
 Matches 1119; Conservative 122; Mismatches 225; Indels 9; Gaps 3;

Qy 1 MONSQDSIFELSLPKGGGATGCGEALTTPGPGMAALSPLPISAGRGYAPAFLLN 60  
 Db 1 MONSQDSIFELSLPKGGGATGCGEALTTPGPGMAALSPLPISAGRGYAPAFLLN 60  
 Qy 61 SGAGNSPFLGWDGCVNMTIRRRTHFGVPHYDETDFTLGPGEVLVVA-----DQPRDES 114  
 Db 61 SGTGNSPFLGWDGCVNMTIRRRTHFGVPHYDETDFTLGPGEVLVVA-----DQPRDES 120  
 Qy 115 TLQGINLGATFTVTGYRSRLESHPFSLRYEQPKTKTDFWLIYSPDQGVHLLKSPQAR 174  
 Db 121 SLQGINLGATFTVTGYRSRLESHPFSLRYEQPKTKTDFWLIYSPDQGVHLLKSPQAR 180  
 Qy 175 ISNPSQTTOTTAQWLLEASVSRGEOIYVYVRAEDDTGCEADEITHLQATQRYLHIVY 234  
 Db 181 ISNPLNVNQAQWLLEASISHSSEIYYQVRAEDAGCETDELAHSAFATVQRLVTHY 240  
 Qy 235 GNRATSETPLGLDGSAPSOADWLFYVDFYGERSNLKTTPAFSTTGSWLCRODRFSRYE 294  
 Db 241 GNLTAADVFTPLNGDDPLKSGWMECLVFDYGERKNSLSEMPLFKATGNWLCRODRFSRYE 300  
 Qy 295 YGFEIRTRRCQVLMYHHLQALDSKITEHNGPTLVSRLLINLYDESAIASTLVFVRVGH 354  
 Db 301 YGFEIRTRRCQVLMYHHLQALDSKITEHNGPTLVSRLLINLYDESAIASTLVFVRVGH 360  
 Qy 355 EQDGNVVTLPPELAYQDFSPRHHQWPMVDVLANFNAIORWQLVDLKGELPGLLYODK 414  
 Db 361 EDNNTVTALPPELAYQDFSPRHHQWPMVDVLANFNAIORWQLVDLKGELPGLLYODK 420  
 Qy 415 GAWWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDQGLDWVITGFLRGY 474

```

421 NGWYRSQAQGBEMNAVTVGWQLLPTTFAVDNLSMDINGDQGLDWITGPGURGY 480
475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAP 534
481 HSQHPDGSWTRFTPLHALPIEYSHPRAQLADLMGAGLSDLVLIQPKSVRLYVNRDGFTE 540
535 GKDVQSGDITLVPVGDPRKLVAFSDVLGSGQAHLEVSATKVTWPNLGRGFRGQPIIT 594
541 GRDVQSGDITLPLPGADARKLVAFSDVLGSGQAHLEVSATQVTCWPNLGHGRFGQPIV 600
595 LPGSQPATEPNPAQVYLADLGGSPDILYVHTNRLDIPLNKGSGAEVPTLFFPEGL 654
601 LPGFSQSAASFPORVHLADLGGSPADLIYVHADRLDIFNSGSGFAKPTLFFPDGL 660
655 RFDHTCOLQMAADVQGLGVASILSVPHMSPHHWRCDLTNMPKPLINENNNNGVHHTLRY 714
661 RPDTCLOQVADVQGLGVVLSLSPVHPMAHHWRCDLTNAKFWLLSETNNNGANHTLHY 720
715 RSSQFWLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTLRYARGMDGR 774
721 RSSQFWLDEKAAALTTGQTPVCYLPPFPVHTLMOTETEDISGNKLVTLRYAGMDGR 780
775 ERFRFGYVQTSQSHQAQNAERTPPALTKWYATGLPIDNALSTYWR-DDQAFA 833
781 ERFRFGYVQTSQSHQAQNAERTPPALTKWYATGLPIDNALSTYWR-DDQAFA 840
834 GFSRFTTQWQNDKVPILTPEDNRYFNENRALKGQLLSELYGLDDSTNKHVPYVTVTEFR 893
841 GFTFRFTLWKGKDVPLTPEDDHNILWLNRAKQGOLSELYGLDGSQQQIPYTVTESR 900
894 SQVRLQHTDSRYFVLSVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQSVQYP 953
901 PQVRLQDQATVSPVLMASVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQSVQYP 960
954 RROQPAINLYPDTLPDKLLANSYDDQQLRLTLVQSSWHHLTNTVTVGLPSTRSDI 1013
961 RNRKPTTNPYPTLTPDLTFASSYDDQQLRLTLVQSSWHHLTNTVTVGLPSTRSDA 1020
1014 FTYGAEVNPAGLNLLELSDKNSLIADKPREYLGQOKTATVTDGQNTPLPTPTQALIA 1073
1021 FTYDAKVVDVGLNLETCAENSLIADKPREYLNQRTFTYDGNQNTPLKPTPTQALIA 1080
1074 FTETVFNQSTLSAFNGSIPSDKLSLTLEAGYQQTNYLFPRTGEDKVVVAHGYTDYGT 1133
1081 FTETAVLTESLSAFDGGITPDELPGILTQAGYQEQEPYLPFTGKENVVARQGVTDYGT 1140
1134 AAQFWRPQKQNTQLTGKILTIWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1193
1141 EAQFWRPVPAQRNLSLTGKWTLLKWDTHYCVITQTQDAAGLTTSYANYDWRFLTPVQLTDIND 1200
1194 NOHLITLDALGRPITLRFWGTENGWTCYSPEKASRSPSDVNAAIELEKPLPVAOCOV 1253
1201 NVHLITLDALGRPVTQRFWGTESGATYSSEKSPFPNDIDTAINLTGFLPVAQCLV 1260
1254 YAPESWMPVLQKTFNRLAEQDWKLVNARIITEDGRICITLAYRRWQSKAIPQLISLL 1313
1261 YAPDSWMPVLFQEFNTLTQEEQETLDRSRIITEDWRICALTRRWLQSKISTPLVKLL 1320
1314 NNGRPLPHSLTTLTDRYDHPDQEQIROOVFPSPDGFGRLLQAAAHHEAGMARQNEGSL 1373
1321 TNSIGLPHNLTLTDRYDRDSEQIROOVAFSDGFGRLLQDSVRHEAGEAWQRNQGSL 1380
1374 IINVOHTENRVAWTCRTEDYKNGQPIRTYQVFLNDWRYVYVNDGARQEAYATHVYDP 1433
1381 VTKVENTKTRVAWTCRTEDYKNGQPIRTYQVFLNDWRYVYVNDGARQEAYATHVYDP 1438
1434 IGRIRVITAKGWRFLTFPFWTFVNEDENTAAE 1468
1439 IGRIRVITAKGWRFLTFPFWTFVNEDENTAAD 1473

```

RESULT 8

```

ABM70526
ID ABM70526 standard; protein; 1486 AA.
XX AC ABM70526;
XX DT 20-NOV-2003 (first entry)
XX DE Photorhabdus luminescens protein sequence #3623.
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX KW whooping cough.
XX OS Photorhabdus luminescens.
XX PN WO200294867-A2.
XX XX 28-NOV-2002.
XX XX 07-FEB-2002; 2002WO-IB003040.
XX XX 07-FEB-2001; 2001FR-00001659.
XX XX (INSP ) INST PASTEUR.
XX XX (CNRS ) CNRS CENT NAT RECH SCI.
XX XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX XX Buchrieser C;
XX XX WPI; 2003-148459/14.
XX XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX XX Claim 2; SEQ ID NO 3623; 1205pp; French.
XX XX The invention relates to the isolation of genes and their encoded
XX XX proteins from Photorhabdus luminescens. The isolated sequences are
XX XX sources of probes and primers for detecting the genome of P. luminescens
XX XX and related species; to study polymorphisms; for gene analysis and for
XX XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX XX polypeptides encoded by the genes are used for detection/identification
XX XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX XX carry a gene-containing vector are used to select compounds that
XX XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX XX animals or microorganisms other than P. luminescens and are able to alter
XX XX response or sensitivity to toxins and antibiotics produced by P.
XX XX luminescens. Cells transformed to express the genes are useful for
XX XX recombinant production of the proteins, particularly toxins and
XX XX antibacterials useful as insecticides, bactericides and fungicides. The
XX XX genes, proteins, vectors containing the genes and Ab are also useful
XX XX therapeutically to treat microbial infection by bacteria or fungi that
XX XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX XX biopesticides. Other uses of the genes and the proteins are as virulence
XX XX factors and for identifying targets of human diseases for which P.
XX XX luminescens is a model (particularly plague and whooping cough). This
XX XX sequence represents one of the isolated P. luminescens proteins
XX XX
XX XX Sequence 1486 AA;

```

```

Query Match 58.6%; Score 4627.5; DB 6; Length 1486;
Best Local Similarity 58.4%; Pred. No. 0;
Matches 869; Conservative 223; Mismatches 368; Indels 29; Gaps 11;

Qy 1 MNSQDSFISLTLSPKGGGATGHEALTPTGPGMAALSPLPISAGRGVAPATLYN 60
Db 1 MQDSFEVSTLTLSPKGGGATGHEALTPTGPGMAALSPLPISAGRGVAPATLYN 60
Qy 61 SGAGNSPFLGWCNVMVTRRTHFGVPHYDETDTFLQPEGEVLVVA-----DQPRDE 113
Db 61 SSAGNGFFGIGWQGVMTISRTHGIPQYGNDDTFLSPQGEVNNIALNQGQPIRDQV 120

```





DR WPI; 1998-179427/16.  
 DR N-ESDB; AAV29224.  
 PT Isolated toxins from Photorhabdus luminescens strains - useful for  
 XX control of insect pests.  
 XX  
 XX Claim 34; Page 196-200; 321pp; English.  
 XX  
 CC The present sequence represents a protein named TcaC of the bacterium  
 CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
 CC nematodes of the Heterorhabdidae genus. The bacterium has at least 4  
 CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
 CC produced from these regions that are associated with insecticidal  
 CC activity. The native toxins are secreted proteins. The proteins are toxic  
 CC to insects upon exposure and especially when ingested. The nucleic acid  
 CC sequence can be used to produce transgenic plants, baculoviruses or  
 CC microbial hosts for toxin production. They can be used to control insects  
 CC pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,  
 CC Dictyoptera, Acarina or Homoptera orders, especially the Southern or  
 CC Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,  
 CC turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,  
 CC corn earworm, European corn borer or tobacco hornworm or budworm  
 XX  
 SQ Sequence 1485 AA;

Query Match 58.0%; Score 4581.5; DB 2; Length 1485;  
 Best Local Similarity 57.6%; Pred. No. 0;  
 Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

QY 1 MONSDRSITSLPKGGGALTGCEALTPTGPDGMAALSPLPISAGRGVAPFTLNYN 60  
 DB 1 MQDSPEVSITLSPKGGGALTNGEALNAGPDGMAALSPLPISAGRGVAPFTLNYN 60  
 QY 61 SGAGNSPFLGWCNVMVTRRTHGPHYDPTDFLPGEGEVLVA-----DQPRDE 113  
 DB 61 NSAGNPPGIGWQCQWMSISRTTGHGIPQYGNDDTFLSPQGEVNMIALNDGQPDQDV 120  
 QY 114 STLOGINLGAFTVGYRSRLSHESFRLHYWPKT--TGKTDWLYSPDGOVHLGKSP 171  
 DB 121 KTLQGVTLPISTVTRYQARQLDPSKLEYWQPAAGSGQGRA-FWLDISTPDGHLHLGKTA 179  
 QY 172 QARISNPSTQTQAQWLEASVSRSGEQIYQYRAEDDTGCEADBITHLQATQRYLHI 231  
 DB 180 QACLANPQNDQIAQWLEEVTPAGEHVSQYRAEDDAHCDNEKTAHPNVAQRYLVQ 239  
 QY 232 VYGNRTASETPLGLDGSAPQADWLFYLVFDYGYRSNNLTPPAFST-TGSWLCQRDRP 290  
 DB 240 VYGNRIKQASLFLVDNAPPAPAEWLFHLPDGHGERTSLHTVPTWADAGTAQWSVRPDI 299  
 QY 291 SRYEYGFIRTRRLCROVLMYHHLQALDLSKITEHNGPTLVSRLLIYNDDESALSTLVFVR 350  
 DB 300 SRYEYGFIRTRRLCROVLMYHHLQALDLSKITEHNGPTLVSRLLIYNDDESALSTLVFVR 359  
 QY 351 RVGHQDGNVTLPLLELAYQDFSPRHAAHQPMQDVLAFNAIQRWQLVDLKGELPGLL 410  
 DB 360 QLSHESDGRPTQPLELAWQRFDEKFTWQRFALDNFNSQRYQLVDLKGELPGLL 419  
 QY 411 YQDGAWYRSARLGEISDAVTEWKMQLSVIPSLQSNASLVDINGDGLDWWITGPG 470  
 DB 420 YQDRGAWYKAPQREGDGSNAVTDKIAPLTFLPNLQDNASLMDINGDGLDWWITGPG 479  
 QY 471 LRGHVSQRPDGSWTFTPLNALPVEYTHPRQAOLADLMGAGLSDLVLIGPKSVRLYANTRD 530  
 DB 480 LRGHVSQRPDGSWTFTPLNALPVEYTHPRQAOLADLMGAGLSDLVLIGPKSVRLYANTRD 539  
 QY 531 GFAKGQVVSQSDITPLPGADPRKLVAFSDVLGSGQAHVVSATKVTCPNLRGRFG 590  
 DB 540 GWRKEDVPQSGITLPTVGTDAKLVAFSDMLGSGQAHVVSATKVTCPNLRGRFG 599  
 QY 591 QPITLPGSQPATEFNPAQVYADLDGSGPTDLIVVHNRDLDFLNKSGNGPABVPLRF 650  
 DB 600 QPITLPGSQPATEFNPAQVYADLDGSGPTDLIVVHNRDLDFLNKSGNGPABVPLRF 659

QY 651 PEGLRFDHTCOLOMADVOGLCVASLILSVPHMSPHHWRCDLTNNKPMWLNENNNMGVHH 710  
 DB 660 PEGVQFNTCOLOVADIQGLGASLILTVPHIAHHWRCDLSLTKPMLNWNNGGAHH 719  
 QY 711 TLYRSSQFWLDEKAAALTTGTPVCYLPPIHTMQTETETDEISGNKLVTTLYARGA 770  
 DB 720 TLYRSSAQFWLDEKQLTRAGKSPACVLPFPMHLLWYETIQDBISGNRLTSEVNSHGV 779  
 QY 771 WDGERFERGVEYQTDHSHOLAQGNAPERTPPALTKNWYATGLPVDINALSTEVNR-DD 829  
 DB 780 WDGERFERGVEYQTDHSHOLAQGNAPERTPPALTKNWYATGLPVDINALSTEVNR-DD 839  
 QY 830 QAFAGSPRFTTQ--DNKDVLPTEPDNSRYFNWNRALKGQLLRSELYGLDDSTNKHVPY 887  
 DB 840 QAFAGSPRFTTQ--DNKDVLPTEPDNSRYFNWNRALKGQLLRSELYGLDDSTNKHVPY 898  
 QY 888 TVTFRFRQVRRLQHTDSRYPVLSVSVESRYHYERIASDPQCQNITLSSDRFQPLKQ 947  
 DB 899 TVSESRYQVRSIPVKNKETELSAWVTAIENRSYHYERIITDPQFSQSIKLOHDIQFQSLQS 958  
 QY 948 LSVQYPRQPAHLYPDTPDKLLANSYDDQORQLRLTYQSSWHHLTNNVTVVLGPD 1007  
 DB 959 VDIAMPREKPAVNPYPTLPELTFDSYDDQQLLRLVROKNSWHHLTNGENWRLGPN 1018  
 QY 1008 STRDIFTYGAENVPAAGLNLLELLSDKNLSIADDPREYLQOQKTAFTYDQNTTPTQPT 1067  
 DB 1019 AQRDVTYDRSKIPTTEGISLEILLKODGLLADEKAAVYLGQOQTFYTAGAEVTLKPT 1078  
 QY 1068 RQALIAFTTFTVFNQSTLSAFNGSIPSDKLTSTLEAGYQQTNYLPPRTGDKVWVAHHG 1127  
 DB 1079 LQALVAFQETAMDDTSLQAYEGVIEQELNALTALQAGYQOVARLENTRSPVMAARQG 1138  
 QY 1128 YTDYGTAAQFWRPQKSNQTLTKITLWDANYCVVOTRDAAGITTSKAYDWRFLTPVQ 1187  
 DB 1139 YTDYGTAAQFWRPQKSNQTLTKITLWDANYCVVOTRDAAGITTSKAYDWRFLTPVQ 1198  
 QY 1188 LTDINDNQHLLTLDALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAAILKPLP 1247  
 DB 1199 LTDINDNQHLLTLDALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAAILKPLP 1255  
 QY 1248 VAQCVVAPESWMPVLSQKTNRL---AEQDWKLYNARIITEDGRICITLAYRWVQSQK 1304  
 DB 1256 VAQCVVAVDSWMPVLSQKTNRL---AEQDWKLYNARIITEDGRICITLAYRWVQSQK 1315  
 QY 1305 AIPOLISLNNPRLPHSLTTLTDYDHPDQQLRQOVVFSDFGRLQLQAAARHAGWA 1364  
 DB 1316 LTIQLISLASIPRLPHSLTTLTDYDHPDQQLRQOVVFSDFGRLQLQAAARHAGWA 1375  
 QY 1365 RQVEDGSLIINVQ-----HTENRWAVTGETYDNKGQPIRTYQYFLNDWRYVNSDA 1418  
 DB 1376 RQVEDGSLIINVQ-----HTENRWAVTGETYDNKGQPIRTYQYFLNDWRYVNSDA 1435  
 QY 1419 ROEKAYADTHVYDPIGREIKVITAKGWFRRLTFTPFWTFVNEDENDTAA 1467  
 DB 1436 RD--DLFADTHLYDPLGREGYKIVITAKKYLREKLYTPFWFVSEDENDTAS 1482

RESULT 11

AAW17887  
 ID AAW17887 standard; protein; 1485 AA.  
 XX  
 AC AAW17887;  
 XX  
 DT 17-OCT-2003 (revised)  
 DT 29-JAN-1998 (first entry)  
 XX  
 DE Photorhabdus luminescens insect toxin protein TcaC.  
 KW Insecticide; insect; toxin; pest control; biological control;  
 KW Photorhabdus luminescens; TcaC; Southern corn rootworm;  
 KW Colorado potato beetle; Western corn rootworm; meal worm; boll weevil;  
 KW turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper;  
 KW codling moth; corn earworm; European corn borer; tobacco hornworm;

KW tobacco budworm; Lepidoptera; Hymenoptera; Diptera, Dictyoptera; Acarina; Homoptera.

XX OS Photorhabdus luminescens; strain W-14 (ATCC 55397).

XX PN WO9717432-A1.

XX PD 15-MAY-1997.

XX PF 06-NOV-1996; 96WO-US018003.

XX PR 06-NOV-1995; 96US-0007255P.

XX PR 28-FEB-1996; 96US-00608423.

XX PR 28-AUG-1996; 96US-00705484.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;

PI Ffrench-Constant RH, Rocheleau TA, Blackburn MB, Hey TD, Merlo DJ;

PI Orr GL, Roberts JL, Strickland JA, Guo L, Ciche T;

XX DR WPI; 1997-281022/25.

XX DR N-PSDB; AAT68840.

XX PT Photorhabdus sp. insecticidal protein toxins and DNA encoding them - can be genetically engineered into insect larvae food and plants for insect control.

XX PS Claim 34; Page 159-164; 276pp; English.

XX CC This polypeptide comprises a specifically claimed insecticidal toxin protein, TcaA, of Photorhabdus luminescens, a 184 kDa component of a toxin protein complex. TcaB can be expressed in host cells using a gene (see AAT68837) isolated from a genomic library. Claimed toxin proteins of P. luminescens (see AAM17871, AAM17884-89, AAM17899-900, AAM18301-06) can be applied to, or genetically engineered into, insect larvae food and plants for insect control. The Photorhabdus toxins are particularly effective against Southern corn rootworm, Colorado potato beetle, Western corn rootworm, meal worm, boll weevil and turf grub (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling moth, corn earworm, European corn borer, tobacco hornworm and tobacco budworm (Lepidoptera), and are also active against insects of the orders Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All claimed). (Updated on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 1485 AA;

Query Match 58.0%; Score 4580.5; DB 2; Length 1485;

Best Local Similarity 57.6%; Pred. No. 0;

Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;

QY 1 MONSDPSITELSLPKGGGAIAGMGALTPGPDGMAALSPLPISAGRGVAPAFPLN 60

DB 1 MODSEVSIITLSLPGGGAINGMGALNAGPDGMAALSPLPISAGRGVAPAFPLN 60

QY 61 SAGNSPFLGMDNCVMITRRTHRGVPHYDETDFTLGPGEGLVVA-----DQPRDE 113

DB 61 NSAGNGPFGIGWQCVMSISRTTGIGIQYGNDDTFLSPQGEVMNIALNDQGPDIRDV 120

QY 114 STLOGINIGATFTVGYRSRLSHESLEYWPKT--TGKTDFTWLYSPDQGVHLLGKSP 171

DB 121 KTLQGVTLPISYTVTRYQARQLDPSKIEYWPASQGEGR-AFWLISTPDGHLILGKTA 179

QY 172 QARISNPQTQTQAQWLEASVSSRGEQIYYOYRAEDDTGCEADSIHTHLQATQRYLHI 231

DB 180 QAQLANPNDQOIAQLLEETVPAGEHVSQYVRAEDDAHCDNDEKTAHPNLTAQRYLVQ 239

QY 232 VYGNRTASETLPGLDGSAQDWFYLVFDYGRSNLKTPTTAPFST-TGSWLCQRDRF 290

DB 240 VYGNIKPKQASLFLVDNAPPAPEEWLFLHVPDHGERDTSHTVPTWDAGTAQWSVRPDI 299

QY 291 SRYEYGFIRTRRLCRQVLMTHLQALDSKTEHNGPTLVSRLLIYNDESALSTLVFR 350

DB 300 SRYEYGFVTRRLCQQVLMFHRFRTALMAGEASTNDAPBLVGRLLILEYDKNASVTLLITIR 359

QY 351 RVGEHQDGNVVTLPPELAYQDFSPRHHARQPMDFLANFNAIQRMQLVDLKEGLPLGL 410

DB 360 QLSHESDGRPVTPQPELEAWQRFLEKTPQWPFALDNFNSQQRVQLVLRGEGPLGWL 419

QY 411 YODKGAWYRSAQRLEIGSDAVTWKQPIVSIPSLQSNASLVINDGQGLDWTGPG 470

DB 420 YQDRGAWYKAPQREQDGSNAVYDKIAPLPTLNLQNASLMDINGDQGLDWTGVTAS 479

QY 471 LRGYHSORPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVIGPKSVRLVYANTRD 530

DB 480 IRGYHSQPDGKWTHTFTFIALPVEYFHPSTQFADLTGAGLSDLVIGPKSVRLVYANQRN 539

QY 531 GFAKGQYVQSGDITLVPFGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLGRGRFG 590

DB 540 GWRKGEDVPQGTGITLPTGTGDKLVAFSDMLSGSQQLHVEIKGNRVTCWPNLGHGRFG 599

QY 591 QPITLPGFSQPATFNPAPQVYADLDGSGPTDLIVHTNRLDIFLUNKSGNGFAEPVTLRF 650

DB 600 QPLTSLGFSQPENGFNPERLEFADIDGSGTDLIYAQSGSLIYLQNSGNQFADPLTAL 659

QY 651 PEGRLFDHTCOLQWADVQGLGVASILSVPHMSPHHRCDLTNMKPWLNNMNNMGVHH 710

DB 660 PEGVQFDNTCOLQVADIQGLGIASLILVPHIAPHHRCDSLTKPWLNVNNNRGAHH 719

QY 711 TLRYRSSQFWLDEKAAALTTGQTPVCYLPFPIHTLMQTEDEISGNKLVTLTYARGA 770

DB 720 TLHYRSSAQFWLDEKQLTKAGKSPACVLPFPMHLLWYTEIQDEISGNRLSEVNSHGV 779

QY 771 WDGEREPRGYVEQTDHSHOLAQGNAPERTPPALTQKNYATGLPVIDNALSTYWR-DD 829

DB 780 WDGEREPRGYVEQTDHSHOLAQGNAPERTPPALTQKNYATGLPVIDNALSTYWR-DD 839

QY 830 QAFAGSPREFTWQ--DNKQVPLPEDDNSRYWFRALKGQLRLSELXGLDSDTNKHVPY 887

DB 840 QAYSGFETRYVWHTNTQDQAFTP-NETQRNLTALKGQLRLSELXGLDSDTNKHVPY 898

QY 888 TVTFERSQVRRLQHTDSRYVPLWSVSVESRNYHERIASDPCCSQNITLSSDRFCQPLKQ 947

DB 899 TVSESRYQVRSI PVNKETELSAWTAIENRSVHYERIITDPQFSQSIKLQHDIFQCSLQS 958

QY 948 LSVQYPRRQQAIVNLYPDITLDPKLLANSYDQOQLRLTYQOQSSWHHLTNTNVRVLGLPD 1007

DB 959 VDIAMPREKPAVNPYPPTLPTLFDSSYDQQLLLRLVROKNSWHHLTDGKNWRLGLPN 1018

QY 1008 STRSDIFTYGAENVPAAGLNLLELSDKNLLADKPEREVLGQOKTAYTDGQNTTLPQPT 1067

DB 1019 AQRDVYTYDRSKIPTEGISLEILLKDDGLLADKAAVYLGQQQFTYTAGQAEVTLKPT 1078

QY 1068 QALIAFTETTVFNQSTLSAFNGSIPSDKLSTTLEQAGYQQTNYLFPRTGEDKVVVAHHG 1127

DB 1079 LQALVAQETAMDDNSIQAEGVIEBEQELNTELTAQYQQVARLFTNRSFSPVWAARQ 1138

QY 1128 YTDYGTAAQFWRPQKQSNLTQTKITLIWDANYCVVQTRDAAGLTTSKAYDWRFLTPVQ 1187

DB 1139 YTDYGTAAQFWRPQKQSNLTQTKITLIWDANYCVVQTRDAAGLTTSKAYDWRFLTPVQ 1198

QY 1188 LTDINDNOHLITLALGRPITLRFWGTENGKMTGYSPSEKASFPSPDVNAIELKPLP 1247

DB 1199 LTDINDNOHLITLALGRVITLRFWGTENGKMTGYSPSEKASFPSPDVNAIELKPLP 1255

QY 1248 VAQCVYAPESWMPVLSOKTFNRL---AEODWKLYNARIITEDGRICLTAYRRVVSQK 1304

DB 1256 VAQCLVAVDWSMPSLSLSQSQEAEALWAQLRAAHMITEDGKVCALSGKRGTSQN 1315

QY 1305 AIPOLISLLNNGPLPHSLTLTDYDHDPEQIRQVVFSDGFRLLQAAARHAGMA 1364

DB 1316 LTIQLISLLASIPRLPHVLGITTDYDSDPQQHQQTQVFSFSDGFRLLQSSARHESDA 1375

QY 1365 RORNEDSLINVO-----HTENRWAVTGTEVDNKGQPIRTYOPYFLNDRWYVNDSDA 1418

DB 1376 WORKEDGELVVDANGVLVSAPTDNRWAVSGTETDYGKQPVRTYQPYFLNDRWYVNDSDA 1435







Db 1133 PQISDVAGSSEKVMVARQYTEYGAAQYRFLIQKSLITGKYLTDWTHYCVVVKTED 1192  
 Qy 1169 AAGLTTSKADYWRFLTPVQLTDINDOHLITDLAGRPITTLRFWGTENGKMTGYSSPEKA 1228  
 Db 1193 GAGMTTQAKYDTRFLPQAQLTDINDOHLITVFNALGOVTSRFRWGTENGKISGYTPESK 1252  
 Qy 1229 SPSPPSDVNAATELKPLPVAQCVYAPESMMPVLSQKTFN---RLAQDQWQKLYNARI 1285  
 Db 1253 PTFVPDTEKALALQPTIPVSCNIVVPDSWMRLPQQSLTQQLKEGETLNNALHRAGVV 1312  
 Qy 1286 TEDGICITLAYERWQSKAIPQ----LISLANNPRLPHSLTLTDRYDHDPPQOIRQ 1341  
 Db 1313 TEDGLICELAYERWTKRQATSSMAVTLQOILAQTPROPHPHMTITTDYSDSDSOQLRQ 1372  
 Qy 1342 OVVFSDGFRLLQAAARHEAGMARQKEDGSLIIN-----YQHTENRWAIVTGRTEYDNK 1395  
 Db 1373 SILVSDGFRVLQSQRHEAGMARQKEDGSLVVDNTGKPVVANTTTTWRVSGRTEYDCK 1432  
 Qy 1396 GQPIRTYQPYFNDWRYVNSDARQEKAYADTHVYDPIGRIRKIVITAKGFWFRRLTFPW 1455  
 Db 1433 GQAIRAYLPYFLNDWRYVNSDARD--DLYADTHFYDPLGREYQVKTAKGFWFRNNFMPW 1490  
 Qy 1456 FTVNEDENDTAAEV 1469  
 Db 1491 FTVNEDENDTAAEL 1504

RESULT 13

ID AEB47812 standard; protein; 1506 AA.  
 AC AEB47812;  
 XX  
 DT 22-SEP-2005 (first entry)  
 DE Native XptB1 (xb).  
 XX  
 KW XptB1; toxin; insect resistance; insecticide.  
 XX  
 OS Xenorhabdus bovienii.  
 XX  
 PN US2005155104-A1.  
 PD 14-JUL-2005.  
 XX  
 PF 23-DEC-2004; 2004US-00020849.  
 XX  
 PR 07-JAN-2004; 2004US-0534893P.  
 XX  
 PA (APEL/) APEL-BIRKHOOLD P C.  
 PA (HEYT/) HEY T D.  
 PA (THOM/) THOMPSON R L.  
 PA (MEAD/) MEADE T.  
 PA (LIZS/) LI Z S.  
 PA (RUSS/) RUSSELL S M.  
 PA (SHEE/) SHEETS J J.  
 PA (LIRA/) LIRA J M.  
 PA (FENC/) FENCIL K J.  
 PA (MITC/) MITCHELL J C.  
 XX  
 PI ApeI-Birkhold PC, Hey TD, Thompson RL, Meade T, Li ZS;  
 PI Russell SM, Sheets JJ, Lira JM, Fencil KJ, Mitchell JC;  
 XX  
 DR WPI; 2005-496874/50.  
 DR N-PSDB; AEB47811, AEB47817, AEB47819.  
 XX  
 PT New isolated protein and encoding nucleic acid having toxin activity  
 PT against an insect, useful for developing new insecticidal toxins that can  
 PT be used to control insects.  
 XX  
 PS Claim 3; SEQ ID NO 2; 54pp; English.  
 XX  
 CC This sequence represents native XptB1 (xb). XptB1 is one of five TC (toxin

CC complex) proteins from Xenorhabdus. XptA1 is a "stand alone" toxin. XptA2  
 CC also has some stand alone toxin activity. XptB1 and XptC1 are the  
 CC xenorhabdus pototoxins that can enhance the activity of either (or  
 CC both) of the XptA toxins. XptD1 has some level of homology with TccB  
 CC (toxin complex c protein B). This protein has toxin activity against an  
 CC insect. The polynucleotide encoding this protein may be used to transform  
 CC a plant cell, and thereby generate a transgenic plant which may be  
 CC ingested by an insect, whose numbers are subsequently controlled. The  
 CC methods and compositions of the present invention are useful for  
 CC developing new insecticidal toxins and other proteins that can be used to  
 CC control insects.  
 XX  
 SQ Sequence 1506 AA;  
 Query Match 53.9%; Score 4261; DB 9; Length 1506;  
 Best Local Similarity 54.1%; Pred. No. 0;  
 Matches 819; Conservative 228; Mismatches 409; Indels 59; Gaps 21;  
 Qy 2 QNSQDSFTELSPKGGGAIQMGGEALPTGPDGMAALSLPLPISAGRGYAFPLINYS 61  
 Db 3 QDSQDMVTQLSLPKGGGAIQMGGDTISNAGPDGMAASLSPPLPISAGRGGAFLNLSYSS 62  
 Qy 62 GAGNSPFLGMDNCVMTIRRTTHFGVPHYDETDTLFGPEGEVLVVA---DOP--RDEST 115  
 Db 63 GAGNSFGIGWQSSTMAISRRTQHGVPOQHGEDTFLCPMGVMAVAVNQSGQDPVKTDX 122  
 Qy 116 LQGINLGATFTVGYRSRLESFSLRYWQPKT-TGKTDFWLIYSPDGQVHLGLKSPQAR 174  
 Db 123 LLGQLPVYTVTRHOPRNIQHFSLKLEYWQPTDVTETPFWLMYSPDQIHIFGKTEQAQ 182  
 Qy 175 ISNPSQTTQATQWLLLEASVSRGEOIYYQYRAEDDTGCEADITHHLQNTAORYLHIVY 234  
 Db 183 IANPAEVSQLAQWLLLEETVTPAGEHIYYQYRAEDDTCDDSEKNAHPNAAQRYLTQVNY 242  
 Qy 235 GNRFTSETPLGLDGSAPSQAQWLFYLVFDYGCERSNNLKTTPAF-STTGSMLCQRQFSRY 293  
 Db 243 GNITPESLLVLKNTPPADNFWLHFLVFDYGERAQEINTVPPFKAPSNWIKRPFRRF 302  
 Qy 294 EYGFETRRLCRQVLMYHHLQAL-DSKITEHNGPTLVSRILNTYDESAIATLVFVRV 352  
 Db 303 EYGFETRRLCQQLMFHRLKSLAGEQIDGEEIPALVARLLSYDNLDSVTTLTAIRQM 362  
 Qy 353 GHEQGNVTVLPPLLEYAYQDPSPRHHAWQPMQVLANFNAIQWQLVDLKGGLPQLLYQ 412  
 Db 363 AYETDALTIALPLEFDYQPFQKQWQEMQPLAGLNAQQYQLVDLYGEGISGLYQ 422  
 Qy 413 DK-GAWWYRSARQLGEIGS-DAVTWEKMQPLSVIPSLQSNASLVINDGQDMDWITGPG 470  
 Db 423 DRPGAWWYQAPIRQKNVEDINAVTYSPIINPLPKIPSQQDRATLMDIDGGHLDWITAGAG 482  
 Qy 471 LRGHYSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLVANTRD 530  
 Db 483 LQGRYSMQPNGEWTHFIPISALPTEYFHPQAQLADLVGAGLSDALIGPSVRLVANDRG 542  
 Qy 531 GFAGKQVQSGDITLVPFGADPRKLVAFSDVILGSGQAHLEVSATKVTWNLGRGRG 590  
 Db 543 NWKAGINWMPDGVNLPFPGDASSLVAFSDMLGSGQQHLEVAQAQSVKCNWNLGHRFG 602  
 Qy 591 QPITLPGFSQAPATEENPAQVVLADLDGSGPTDILVYHTNRLDILFNKSGNGFAEPVTLRF 650  
 Db 603 AAILPGFSQPNGTFTNANQVFLADLDGSGTADIIYAHSTYLDIYLNESNRRFSAPVRLNL 662  
 Qy 651 PEGRLFDHTCOLQADVQGLGVASLISVPHMSPHHWRCDLTNMKPLNEMNNMGVHH 710  
 Db 663 PEGVNFDTNLCQVSDIOGLGAASIVLTVPHMTPRHWYDFTHKFWLLNVNNRGAET 722  
 Qy 711 TLRYSSSQFWLDEKAAALTTGQTPVCYLPFPPIHTLMQTEDEISGNKLVTLTRYARGA 770  
 Db 723 TLFYSSAQFWLDEKSIIEELKFAASVLPFPPIHLLWRNEALDEITGNRLTKVMYAHGA 782  
 Qy 771 WDGRERERFGYVEQTSQHLAQGNAPERTP-----PALTKWYATGLPVDNALSTSEY 825  
 Db 783 WDGRERERFGYVQIDTDEFAKGT-EKAPDENIYPSRSISWFAATGLFEVDSQLPAEY 841

826 WR-DDQAFAGSPRTTW-----ODNKDVPLTPEDDNSRYWENRALKGQLRLSELXGLD 878  
 842 WRGDDQAFAGTPRTRTRKEKNAGQEGQDTPKEPTETAYWLNAMKQLRSEVYG-D 900  
 879 DSTNK-HVPYVTVTFRSQVRRLQHTDSRYPVLWSVBSRNYHYERIASDPQCQNITLS 937  
 901 DKTEKAKIPYVTEARCQVRLIPSNDEAAPSSWTSIIENRSYHYERIVVDPSCKQOVVIX 960  
 938 SDRFGQPKQLSQVQPRRQQAIPNLYPDTLDPKLLANSYDDQOQRLRLTYQOOSNHHLTN 997  
 961 ADEYGFPLAKVDIAYPRENKPAQNPYDPSLPTDLFADSYDDQOQQLYLTQOQSYHLTQ 1020  
 998 NTVVLGLPDSTRSDIFTY----GAENVPAGGLNELLSDKNSLIADDPKREYLGQOKTA 1053  
 1021 QDDWVLGLTDSRYSEVHYAQTDAQSDIPKAGLILEDLLKVDGLIGDKDTIYLGQOKVA 1080  
 1054 YTDGQNTPLQTPRQALIAETETTVFNQSTLSAFNGSIPSDKLSLTILEQAQYQOTNYLF 1113  
 1081 YVGSD----AEKPTQVRVAYTETAAPDDNALHAFDGVIAPELTQQLAGGY----LLV 1132  
 1114 PRTEG-----DKVVAHHGYTDYGTAAQFWRPQKQSNQTLTKGKITLIWANYCVVQTRD 1168  
 1133 POISDVAGSEKVVVARQGYTEYGAQFYRPLIQKSLLGKVTLSNDTHYCVVVKTED 1192  
 1169 AAGLTTSKDYWRFLTPVOLTINDNOHLITLDALGRPITLRFWGTENGKWTGYSSPEKA 1228  
 1193 GAGMTTQAKYDYLPLPAQLTINDNOHIVTFNALGQVTSRFRWGTENGKISGYSTPEK 1252  
 1229 SFSPPSDVNAALIEKLPVPAQCOVYAPESWMPVLSQTFN---RLABQDWQKLYNARI 1285  
 1253 PFTVPDTEKALALQPTIPVSQCHYYPDSWNRLLPQOSLTGQLKEGETLWNLHRAGV 1312  
 1286 TEDGRICTLAYRWVQSKAIPQ----LISLNNGPRLPPHSLTLTDRYDHDPEQOIRQ 1341  
 1313 TEDGLICELAYRWIKRQATSSMAVTLQOILAQTTPROPHPAMITITDRYDSDSQQLRQ 1372  
 1342 QVPSDGFGRLLQAAAEHAGWABORNEGDSLIIN-----VQHTENRWAVTGRTYDNK 1395  
 1373 SIVLSDGFRVLQSAQRHEAGWABORNEGDSLVVDNTGKPVVANTRWAVSGRTYDVGK 1432  
 1396 QPRTYQYFPLNDWRYVNSDSARQEKEYADTHYDPIGRKEIVITAKGWFRTLFTPW 1455  
 1433 GOAIRAYLPYLNDWRYVNSDSARD--DLXADTHFYDPLGREYQVKTAKGFWRENMFMPW 1490  
 1456 FTVNEDNDTAAEV 1469  
 1491 FVWNEDNDTAAEL 1504

RESULT 14  
 AAY97695  
 ID AAY97695 standard; protein; 1428 AA.  
 XX  
 AC AAY97695;  
 XX  
 DT 19-JUN-2001 (first entry)  
 XX  
 SepB protein encoded by Serratia insecticidal protein complex gene.  
 DE  
 DE Insecticidal protein complex; amber disease; insect; Coleoptera;  
 KW pesticide; SepB protein.  
 KW  
 KW Serratia sp.  
 OS  
 OS WO200116305-A2.  
 FN  
 FN  
 PD 08-MAR-2001.  
 XX  
 XX 04-SEP-2000; 2000WO-NZ000174.  
 XX  
 XX 02-SEP-1999; 99NZ-00337610.  
 PR  
 XX

(NZPA-) NEW ZEALAND PASTORAL AGRIC RES INST LTD.  
 Glare TR, Hurst MRH, Jackson TA;  
 WPI; 2001-169009/17.  
 DR N-PSDB; AAA91292.  
 New nucleic acid encoding a polypeptide useful as a pesticide especially for Coleoptera.  
 Claim 24; Page 100-105; 109pp; English.  
 This sequence represents the SepB protein encoded by the Serratia insecticidal protein complex gene of the invention. The invention relates to a gene encoding an insecticidal protein complex or a functional fragment, a neutral mutation, or a homologue of the complex. The polypeptides and nucleotides of the invention are used to induce amber diseases or like conditions in insects, especially to those from the order comprising Coleoptera, useful as a pesticide  
 Sequence 1428 AA;  
 Query Match 50.3%; Score 3971; DB 4; Length 1428;  
 Best Local Similarity 52.7%; Pred. No. 2.3e-309;  
 Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;  
 QY 1 MONSODESITELSLPKGGGALTGTGGEALTPGPDGMAALSPLPISAGRGVAPFTLNYN 60  
 DB 1 MONHQDMAITAPTLPSCGGGAVTGLKGDIAAAGPDGAATLSIPLVSPGSGVAPFTGALNYH 60  
 QY 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDFTLGPGEVVLWA-----DQPRDES 114  
 DB 61 SRSGNGPFGIGWGIGGAHVQRTRNGAPTVDTDFTGPDGEVLVPALTAAGTOEARQAT 120  
 QY 115 TLQINGLGAFTVTGYSRLESFHSRLRYQPKTKTKDFWLIYSPGQVHLLGKSPQAR 174  
 DB 121 SLLGINPFGSFNVQVYRSRTEGSLRSLERMLPADETEFEWLVYTPDQVALLGRNAQAR 180  
 QY 175 ISNPSQTTQTAQWLEASVSRGEOIYVOYRAEDDTGCEADEITHLQATAQRYLHVY 234  
 DB 181 ISNTAPTQTAWLMESVSLTGEOYIYQRAEDDDGCEAERDAHPQAGRPVAVVY 240  
 QY 235 GNRTASETLPGLDGSAPOADWLFYLVFDYGERSNLKTTPAFSTTGS--WLCRQDRFSR 292  
 DB 241 GNRQAATLPAL-VSTFSMDSWLFILYFDYGERSSVSEAPAWOTPGSGEWLQCQDFSG 299  
 QY 293 YBYGFEIRTRRLCRQVLMYHHLQALDSKI TEHNQPTLVSRILINYDESAIASTLVFRRV 352  
 DB 300 YEFGNLRTRLCRQVLMYHVLGVLAGSSGANDAPALISRLLLDYRESPSLSLENVHQV 359  
 QY 353 GHEODGNVVTLPPLLAYQDFSPRHHAHQPMQMDVLNFAIQRWQLVDLKGEGPGLLYQ 412  
 DB 360 AYESDGTGSCALPALACWQTFPTPLSAWQTRDMDGKLSGLQPYQLVDLNGEGVVGILYQ 419  
 QY 413 DKGAWYRSARQLGEIGSDAVTEWKMQLSVIPSLQSNASLVINDGQDLMDWITGGLR 472  
 DB 420 DSGAWTRREPVRQSGDDPDATWCAAAALPTMPALHNSGILADLNGDGRLEWVTAQVA 479  
 QY 473 GYHSQRPDGSGWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGF 532  
 DB 480 GMYDRTPGRDWLHFTPLSALPVEYAHKPAVLADILGAGLTDMVLIGPVSRLYSGKNDGW 539  
 QY 533 AKGKDVQSGDITLVPFGADPRKLAESDVLGSGQAHLEVSATKVTCPNLRGRFGQP 592  
 DB 540 NKGETVQOTERLTLVPFGVDPRTLVAFSDMAGSQOHLTEVRANGRYWPNLGRFGQP 599  
 QY 593 ITLPFGSQPATEFNPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRPE 652  
 DB 600 VNIFFGSQVTTFNPDQILLADTDGSGTDLIYAMSRLVIYFNQSGNYFAEPHTLLPK 659  
 QY 653 GLRPDHTCQLQMAVQGLGVASLILSVPHMSPHWRCDLTNMKPLLNMNNMNGVHTL 712  
 DB 660 GVRDRTCSLQVADIQGLGVPSLLLTVPVAPHFWCHLSADKDWLLNGMNNMGARHAL 719



473	QY	GYHSQRDPDGSWTRTPFNALPVEYTHPRPAQIADLMGAGLSDLVLIGPKSVRLYANTRDGF	532
480	Db	GMYDRTFGRDNLHFTPLSALPVEYAHKPAVLADTLGAGLTDMVLIGPRSVRLYSGKNDGW	539
533	QY	AKGKDVQSGDITLPPVGAOPKLVAFSDVLGSGOAHLEVSATKVTWCNPLNLRGRFGQP	592
540	Db	NKGETVQOTERLTLPVPGVDPRTLVAFSDMAGSGQHLTEVRANGVRYWNLGHRGQP	599
593	QY	ITLPGFSQAPATERNPAQVYLADLGDGSPDTLLIYVHTNRDLIFLNKSNGFAEPVTLRFFE	652
600	Db	VNIPGFSQSVTTNPNDQILLADTDGSGTDLIYAMSRLVIYFNQSGNYFAEPHTLLLPK	659
653	QY	GLRFDHTCQLOMADVQGLGVASLILSVPHMSPHHRCDLTNMPKFWLLNWNMMGVHHTL	712
660	Db	GVRYDRTCSLQVADIQGLGVPSLLITVPHVAPHHVCHLSADKFPWLLGNWNNMGARHAL	719
713	QY	RYRSSQFWLDEKAAALTTGOTPCVCLPFPHTLWOTETDETSGNKLVTTLRYARGAWD	772
720	Db	HYRSSVQWLDEKAAALAAAGSPACYLPFTLHLWRSVVOQDETGNKLVSDVLYRHGVWD	779
773	QY	GRERPRGFGYEQTDSHQLA-QGNAPERTPPALTKNYYATGLPVIDNALSTEYWRDD-Q	830
780	Db	QSERPRGFGFVEIRDTDTLASQGTATELSMPSVSRNYYATGVPVDERLPETYWQNDAA	839
831	QY	AFAGFSRPFRTWQNDKDVLTPEDDNSRYWFNRALKQIARSELXGLDDSTNKNHPVTVT	890
840	Db	AFADFATRFVSGSEDTYTP-DDSKTFWLQRLKGLILRSELXGLADGSSQADIPVSVT	898
891	QY	EFRSQVRLQHTDSRYVPLWSSVVESNRYHYERIASDPQCSQNTLSSDSFGOPLKOLSV	950
899	Db	ESRPQV-RLVEANGDIYVWVPMGAESITSVYERVHNDPQCCQAVLLSDEYFPLRQVSV	957
951	QY	QYPRRQOAPINLYPDTLPDKLLANSYDDQQRQLRTYQQSSMHHLTNNTVR--VLGLPDS	1008
958	Db	NYPRRPPSADNPYPASLPATLFA NSYDEQQIILRLGLQQSSAHHLVSLSEGHLLGLAEA	1017
1009	QY	TRSDIFYGAE NYPAGLNLLELLSDKNLSLADDKPREVLYGQOKTAYTDGQNTPTLOTPTR	1068
1018	Db	SRDDVFTYSADNPPEGGLTLEHLLAPBSLSQGVGLAGQQVWYLDSDQVATVAAPPL	1077
1069	QY	QALIAFTETTVFNQSTLSAFNGSIPSKLSTLLEQAGYQOOTNLYFP--RTGEDKVVVAHH	1126
1078	Db	PPKVAFTETAVLDEGMVSSLAAYIVDEH---LEQAGYRQSGVLYFPRGREAEQALWTQCQ	1133
1127	QY	GYTDYGTAAQFWRPKQOSNTQLTKGTLIWDANYCVVVQTRDRAAGLTTSKAYDWRFLTPV	1186
1134	Db	GYVTYAGAEHFWLPLSPRDSMLTGPVTVTRDAVDCVITQWQAAGIVTTADYDWRFLTPV	1193
1187	QY	QLTDINDNQHLITLDALGRPTTLRFWGTENGKMTGYSSPEKASFPSPSDVNAALTELKPL	1246
1194	Db	RVTDPNUNLQSVTLDALGRVTTTLRFWGTENGATGYSD---ATLSVPDGAALALTAPL	1250
1247	QY	PVAQCQVYAPESMMPVLVSQKTFNRLABEQDWQKLYNARIITEDGRCTLAYRWVQSOKAI	1306
1251	Db	PVAQCLVYVTDSW-----GDDNDEK-----	1270
1307	QY	PQLISLLNNGRPLPPHSLTLTDRYDHPQOIRQQVVFSDGFGRLLOAAARHEAGMARQ	1366
1271	Db	-----MPPHVVLATDRYDSDTQGVQVQVTFSDGFGRELOASATROAEGNAWQ	1318
1367	QY	RNEDGSLI-----INVQHTENRWAVTGREYDNKQOPRTTQPYFLNDWRVYVNSDSAR	1419
1319	Db	RGRDGLVTASDGLPVTV-AINFRAWVTGRAEYDNKGLPVRYQPYFLDSQVYVSDDSAR	1377
1420	QY	QEKAYADTHVYDPIGREIKVITAKGWFRRFTLFTFPWFTVNEDENDT	1465
1378	Db	Q--DLYADTHFYDPTAREWQVITAKGRRQVLYTFPFWFVSEDENDT	1421

Search completed: February 16, 2006, 21:31:28  
Job time : 154.155 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 133.704 Seconds  
(without alignments)  
4606.314 Million cell updates/sec

Title: US-10-754-115-45  
Perfect score: 7901  
Sequence: 1 MNSQDPSITSLSPKGGGA.....WFTVNEEDNTAEVKKVKM 1474

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA\_Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	4	US-10-706-424-10
2	7901	100.0	1474	4	US-10-754-115-45
3	6031.5	76.3	1476	3	US-09-817-514A-4
4	4581.5	58.0	1485	4	US-10-262-794A-32
5	4261	53.9	1506	4	US-10-754-115-49
6	4261	53.9	1506	6	US-11-020-848-2
7	3971	50.3	1428	4	US-10-754-115-60
8	3834	48.5	1493	4	US-10-753-901-18
9	3834	48.5	1493	4	US-10-754-115-18
10	3106.5	39.3	1444	4	US-10-609-113-11
11	3106.5	39.3	1444	4	US-10-754-115-40
12	1262	16.0	697	4	US-10-609-113-39
13	202	2.6	2386	4	US-10-156-761-7751
14	199	2.5	2060	4	US-10-381-596A-2
15	181.5	2.3	2364	4	US-10-156-761-7834
16	162.5	2.1	1250	4	US-10-156-761-7572
17	160.5	2.0	1385	4	US-10-283-122A-68242
18	160.5	2.0	1426	3	US-09-912-020-340
19	160.5	2.0	1426	4	US-10-282-122A-42617
20	160.5	2.0	1426	5	US-10-771-241-340
21	160.5	2.0	1551	4	US-10-437-963-181412
22	159.5	2.0	1329	5	US-10-450-763-56139
23	159.5	2.0	1329	5	US-10-450-763-58761
24	154	1.9	5215	3	US-09-861-289-2
25	154	1.9	5215	3	US-09-860-846-2
26	154	1.9	5215	3	US-09-988-384B-2
27	154	1.9	5215	3	US-09-836-821-2

## ALIGNMENTS

### RESULT 1

US-10-706-424-10  
; Sequence 10, Application US/10706424  
; Publication No. US20040103455A1  
; GENERAL INFORMATION:  
; APPLICANT: ffrrench-Constant, Richard  
; APPLICANT: Waterfield, Nicholas  
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens  
; FILE REFERENCE: 62878  
; CURRENT APPLICATION NUMBER: US/10/706.424  
; CURRENT FILING DATE: 2003-11-12  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 10  
; LENGTH: 1474  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-10-706-424-10

Query Match 100.0%; Score 7901; DB 4; Length 1474;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNSQDPSITSLSPKGGGALTGMGEALTPTGPDGMAALSLPLPISAGRGVAPAFLLVYN	60
Db	1	MNSQDPSITSLSPKGGGALTGMGEALTPTGPDGMAALSLPLPISAGRGVAPAFLLVYN	60
Qy	61	SGAGNSPFLGWCDCNVMTIRRRTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN	120
Db	61	SGAGNSPFLGWCDCNVMTIRRRTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN	120
Qy	121	LGATFTVTGYSRSLSHSFSLRYWQPTTKTDFWLIYSPDQVHLLKSPQARISNPSQ	180
Db	121	LGATFTVTGYSRSLSHSFSLRYWQPTTKTDFWLIYSPDQVHLLKSPQARISNPSQ	180
Qy	181	TTQTQAWLLEASVSRGEQIYQVRAEDDTGCEADEITHLQATQRYLHVYVGNRTAS	240
Db	181	TTQTQAWLLEASVSRGEQIYQVRAEDDTGCEADEITHLQATQRYLHVYVGNRTAS	240
Qy	241	ETLPGLDGSAPSQADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEYGFIR	300
Db	241	ETLPGLDGSAPSQADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEYGFIR	300
Qy	301	TRRLCRQVLMTHHLQALDSKITEHNGPTLVSRLLIYNDESAIASTLVFVRVGHQDQGNV	360
Db	301	TRRLCRQVLMTHHLQALDSKITEHNGPTLVSRLLIYNDESAIASTLVFVRVGHQDQGNV	360
Qy	361	VTLPPLELAYODFSRRHHAHQPMQMDVLNFAIQRWQLVDLKGEGLLYQDKGAWYR	420
Db	361	VTLPPLELAYODFSRRHHAHQPMQMDVLNFAIQRWQLVDLKGEGLLYQDKGAWYR	420

```

Qy 421 SAQRLGIGSDAVTWKMQPLSVIPSLQSNASLVINDGCDGOLDWVITGPGLRGVHSORPD 480
Db 421 SAQRLGIGSDAVTWKMQPLSVIPSLQSNASLVINDGCDGOLDWVITGPGLRGVHSORPD 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Qy 541 SGDTITLPPVGPADPRKLVAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQPIITLPGFSQ 600
Db 541 SGDTITLPPVGPADPRKLVAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQPIITLPGFSQ 600
Qy 601 PATEFNAQVYADLDGSGPTDLIVHTNRLDIFLNKSGNGFASPVTLRPEGLRFDHTC 660
Db 601 PATEFNAQVYADLDGSGPTDLIVHTNRLDIFLNKSGNGFASPVTLRPEGLRFDHTC 660
Qy 661 QLOWADYQGLGVASLILSVPHMSPHWRCDLTNNKPMLLNEMNNMNGVHHTLRYRSSQF 720
Db 661 QLOWADYQGLGVASLILSVPHMSPHWRCDLTNNKPMLLNEMNNMNGVHHTLRYRSSQF 720
Qy 721 WLDEKAAALTGTQTPVCVLPPIHTLWQTEDEISGNKLVTTILRYARGAWDGRERFRG 780
Db 721 WLDEKAAALTGTQTPVCVLPPIHTLWQTEDEISGNKLVTTILRYARGAWDGRERFRG 780
Qy 781 FGYVEQTDHQAQGNAPERTPPALTKNWTATGLPVIDNALSTYEWDDQAFAGFSRPT 840
Db 781 FGYVEQTDHQAQGNAPERTPPALTKNWTATGLPVIDNALSTYEWDDQAFAGFSRPT 840
Qy 841 TWQNDKDVLTPTDNDNRYWPNRALKQOLLSELYGLDDSTNKHPVTVTFEFSQVRRLQ 900
Db 841 TWQNDKDVLTPTDNDNRYWPNRALKQOLLSELYGLDDSTNKHPVTVTFEFSQVRRLQ 900
Qy 901 HTDSRYVPLMSSVVESSRYHYERIASDPQCSQNTILSSDRFGQPLKQLSVQYPRRQPAI 960
Db 901 HTDSRYVPLMSSVVESSRYHYERIASDPQCSQNTILSSDRFGQPLKQLSVQYPRRQPAI 960
Qy 961 NLVYEDTLTDLKANSYDDQORQLTYQSSWHLTNNTVRVGLPDSSTRSDIPTFYGAEN 1020
Db 961 NLVYEDTLTDLKANSYDDQORQLTYQSSWHLTNNTVRVGLPDSSTRSDIPTFYGAEN 1020
Qy 1021 VPAGLNLLELSDKNSLIADKPKEYLGQQKTAITDQGNTPPTQTPRQALIAFTETTVF 1080
Db 1021 VPAGLNLLELSDKNSLIADKPKEYLGQQKTAITDQGNTPPTQTPRQALIAFTETTVF 1080
Qy 1081 NQSTLSAFNGSIPSDKLSSTLLEQAGYQNTNLPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
Db 1081 NQSTLSAFNGSIPSDKLSSTLLEQAGYQNTNLPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
Qy 1141 QKQNTQLTGKILTIWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDINDNQHILITL 1200
Db 1141 QKQNTQLTGKILTIWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDINDNQHILITL 1200
Qy 1201 DALGRPITLRFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPSSWM 1260
Db 1201 DALGRPITLRFWGTENGKMTGYSSPEKASFPSPDVNAAIELKKPLPVAQCQVYAPSSWM 1260
Qy 1261 PVLQKTFNRLAEDQWKLNNARIITEDGRICTLAYRRVQSQKAIQPLISLLNNGPRLP 1320
Db 1261 PVLQKTFNRLAEDQWKLNNARIITEDGRICTLAYRRVQSQKAIQPLISLLNNGPRLP 1320
Qy 1321 PHSITLTTDRYDHDPEQIIRQOVVFSQGRGLLOAAARHAGMARQORNEGSLIINVQHT 1380
Db 1321 PHSITLTTDRYDHDPEQIIRQOVVFSQGRGLLOAAARHAGMARQORNEGSLIINVQHT 1380
Qy 1381 ENRWAVTGRTEYDNKGQPIRTYQYFNLNDRYVNSDSARQEKAYADTHVYDPTIGREIKV 1440
Db 1381 ENRWAVTGRTEYDNKGQPIRTYQYFNLNDRYVNSDSARQEKAYADTHVYDPTIGREIKV 1440
Qy 1441 ITAKGNFRRTLTPTPWFTVNEDENDTAAEYKVKVM 1474
Db 1441 ITAKGNFRRTLTPTPWFTVNEDENDTAAEYKVKVM 1474

```

```

RESULT 2
US-10-754-115-45
; Sequence 45, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 1474
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-45

Query Match 100.0%; Score 7901; DB 4; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNSQDPSITELSLPKGGGAIQMGEGALTPGPDGMAALSPLPISAGRGYAPFTLNYN 60
Db 1 MNSQDPSITELSLPKGGGAIQMGEGALTPGPDGMAALSPLPISAGRGYAPFTLNYN 60
Qy 61 SGAGNSPFLGWCDCNVTIRRRTHFGVPHYDETDFTFLGPEGEVLVWADQPRDESTLQGIN 120
Db 61 SGAGNSPFLGWCDCNVTIRRRTHFGVPHYDETDFTFLGPEGEVLVWADQPRDESTLQGIN 120
Qy 121 LGATFTVTGYSRLESFHSRLEYWQPKTKTDFWLIYSPDGOVHLLGKSPQARISNPSQ 180
Db 121 LGATFTVTGYSRLESFHSRLEYWQPKTKTDFWLIYSPDGOVHLLGKSPQARISNPSQ 180
Qy 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVYGNRTAS 240
Db 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVYGNRTAS 240
Qy 241 ETLPGLDGSAPSQADWLFLYLVFDYGERSNMLKTPPAFTTGSWLCRODRFSRYEYGEIR 300
Db 241 ETLPGLDGSAPSQADWLFLYLVFDYGERSNMLKTPPAFTTGSWLCRODRFSRYEYGEIR 300
Qy 301 TRRLCROVLVYHHLQALDSKITEHNGPTLVSRLLINLYNDESALASTLVFVRVHGHEQGNV 360
Db 301 TRRLCROVLVYHHLQALDSKITEHNGPTLVSRLLINLYNDESALASTLVFVRVHGHEQGNV 360
Qy 361 VTLPPLLEAYQDPSPRHHAHQMDVLANFNAIQRWOLVDLKGEGFLGLLYQDKGAWMYR 420
Db 361 VTLPPLLEAYQDPSPRHHAHQMDVLANFNAIQRWOLVDLKGEGFLGLLYQDKGAWMYR 420
Qy 421 SAQRLGIGSDAVTWKMQPLSVIPSLQSNASLVINDGCDGOLDWVITGPGLRGVHSORPD 480
Db 421 SAQRLGIGSDAVTWKMQPLSVIPSLQSNASLVINDGCDGOLDWVITGPGLRGVHSORPD 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Qy 541 SGDTITLPPVGPADPRKLVAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQPIITLPGFSQ 600
Db 541 SGDTITLPPVGPADPRKLVAFSDVLGSGQAHLEVESATKVTWPNLGRGRFGQPIITLPGFSQ 600

```

```

Db 541 SGGDITLVPFGADPRKLVAFSDVLSGQAHLEVSATKVTCPNPLGRGRFGQPIITLPGFSQ 600
Qy 601 PATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGLAFDHTC 660
Db 601 PATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGLAFDHTC 660
Qy 661 QLOMADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNGVHTLRYRSSQF 720
Db 661 QLOMADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNGVHTLRYRSSQF 720
Qy 721 WLDEKAAALTTGQTPVCLPPIHTLWOTETEDISGNKLVTLRYARGAMDGRERFRG 780
Db 721 WLDEKAAALTTGQTPVCLPPIHTLWOTETEDISGNKLVTLRYARGAMDGRERFRG 780
Qy 781 FGYYEQTDSSHQAQNAPEPTPALTKWYATGLPVIDNALSTEWYRDDQAFAGSPRFT 840
Db 781 FGYYEQTDSSHQAQNAPEPTPALTKWYATGLPVIDNALSTEWYRDDQAFAGSPRFT 840
Qy 841 TWQDNKDVPLTPEDDNSRYWFRNALKGQLLRSELYGLDDSTNKHVPYVTFERSQVRRLQ 900
Db 841 TWQDNKDVPLTPEDDNSRYWFRNALKGQLLRSELYGLDDSTNKHVPYVTFERSQVRRLQ 900
Qy 901 HTDSRYPVLMSSVVSERNYHRIASDPQSONITLSSDRFCQPLKQLSVQVPRQQPAI 960
Db 901 HTDSRYPVLMSSVVSERNYHRIASDPQSONITLSSDRFCQPLKQLSVQVPRQQPAI 960
Qy 961 NLYPDTLPDKLLANSYDDQQRQLRTYQOSSWHHLTNNTVRVLGLPDSRSDIFTYGAEN 1020
Db 961 NLYPDTLPDKLLANSYDDQQRQLRTYQOSSWHHLTNNTVRVLGLPDSRSDIFTYGAEN 1020
Qy 1021 VPAGGLNLELLSDKNSLIADDPKRYLGOQKATAYTDGQNTTPLOTPTRQALIAFTETTVF 1080
Db 1021 VPAGGLNLELLSDKNSLIADDPKRYLGOQKATAYTDGQNTTPLOTPTRQALIAFTETTVF 1080
Qy 1081 NOSTLSAFNGSLPSDKLSTLEAQGYQQTNYLFPRTGEDKVVAHHGYTDYGTAAQFWRP 1140
Db 1081 NOSTLSAFNGSLPSDKLSTLEAQGYQQTNYLFPRTGEDKVVAHHGYTDYGTAAQFWRP 1140
Qy 1141 QKQNTQLTQKITLWDANYCVVQTRDAAGLTTSAKYDWRFLTPVQLTDINDNQHLLITL 1200
Db 1141 QKQNTQLTQKITLWDANYCVVQTRDAAGLTTSAKYDWRFLTPVQLTDINDNQHLLITL 1200
Qy 1201 DALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAALIELKPLPVAQCQVAPESWM 1260
Db 1201 DALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAALIELKPLPVAQCQVAPESWM 1260
Qy 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRICTLAYRWVOSOKAIPLIILNNGPRLP 1320
Db 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRICTLAYRWVOSOKAIPLIILNNGPRLP 1320
Qy 1321 PHSLTLTTRDYDHPDPEQIRQQVFSDFGRLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380
Db 1321 PHSLTLTTRDYDHPDPEQIRQQVFSDFGRLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380
Qy 1381 ENRWAVTORTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVDPDGREIKV 1440
Db 1381 ENRWAVTORTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVDPDGREIKV 1440
Qy 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAAEVKVKVM 1474
Db 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAAEVKVKVM 1474

```

RESULT 3

```

US-09-817-514A-4
; Sequence 4, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

```

```

; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1476
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-4

Query Match 76.3%; Score 6031.5; DB 3; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Qy 1 MONQDSFISITELSLPKGGGALTGMGEALTPTGPGMAALSPLPISAGRGYAPAFITLNVN 60
Db 1 MONQTFVTELSUPKGGGALTGMGEALTPTGPGMAALSPLPISAGRGYAPAFITLNVN 60
Qy 61 SGAGNSPFGLGWDCNVMITRRTHFGVPHYDETDTFLGPEGEVLVVA-----DOPRDS 114
Db 61 SGTGNSPFGLGWDCGWAIRRTSTGVVNYDETDTFLGPEGEVLVVALNEAGQADIRSS 120
Qy 115 TLOGINLGATPTVTGYRSLRSLRLEYWQPKTKTDFWLIYSPDGQVHLLGKSPQAR 174
Db 121 SLQGINLGATPTVTGYRSLRSLRLEYWQPKTKTDFWLIYSPDGQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQTAQMLLEASVSSRGQIYYQYRAEDDTGCEADEITHHQAQYRHLHYVY 234
Db 181 ISNPLNVNQTAAQMLLEASISHSBQIYYQYRAEDDEAGCETDELAHAHSAFVQYIQTVHY 240
Qy 235 GNRITASEPLTGLDGSAPSQADWLFLYFVGYGERSNNLKTTPAFSTTGSWLCRQDRFSRYE 294
Db 241 GNLITASDVFPTLNGDDPLKSGMFCFLVFDYGERKNSLSEMLPFKATGNWLCRDRFSRYE 300
Qy 295 YGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAIASTLVFVRVGH 354
Db 301 YGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAIASTLVFVRVGH 360
Qy 355 EQDGNVVTLPLELAYQDPSRHHAHQPMQVLANFNAIQWQVLDLAGEGLPGLLYQDK 414
Db 361 EDNNTVTALPPELAYQDPEPEQATLWQMDVLANFNTIQWQLDLAGEVPGVGLYQDR 420
Qy 415 GAWYRSQRLGEIGSDAVTWKMQPLSVIPSLOSASLVINDINGDQDLWDWITGFLRGY 474
Db 421 NGWYRSQRLGEIGSDAVTWKMQPLSVIPSLOSASLVINDINGDQDLWDWITGFLRGY 480
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVIGPKSVRLYANTRDGFAK 534
Db 481 HSQHPDGSWTRFTPLHALPIEYSHPRQAQLADLMGAGLSDLVIGPKSVRLYANNRDGETE 540
Qy 535 GKDVVQSGDITLVPFGADPRKLVAFSDVLSGQAHLEVSATKVTCPNPLGRGRFGQPIIT 594
Db 541 GRDVVQSGDITLVPFGADPRKLVAFSDVLSGQAHLEVSATKVTCPNPLGRGRFGQPIIV 600
Qy 595 LPGFSQAPATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGL 654
Db 601 LPGFSQSAASFNPDRLVHLADLGGSPADLIYVHADRLDI FSNESGNGFAKPTLSFPDGL 660
Qy 655 RFDHTCQLQOMADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNGVHTLRY 714
Db 661 RFDHTCQLQVADVQGLGVVSLILSVPHMAPHHRCDLTNAPKFWLLSETNNMNGVHTLRY 720
Qy 715 RSSQFNLDEKAAALTTGQTPVCLPPIHTLWOTETEDISGNKLVTLRYARGAMDGR 774
Db 721 RSSQFNLDEKAAALATGQTPVCLPPIPVHTLWOTETEDISGNKLVTLRYARGAMDGR 780
Qy 775 EREPRFGYVEQTDSSHQAQNAPEPTPALTKWYATGLPVIDNALSTEWYRDDQAF 833
Db 781 EREPRFGYVEQTDSSHQAQNAPEPTPALTKWYATGLPAVDNALSAGYWRGDKQAF 840

```

```

Qy 834 GFSRFTTQNDKDVLPDTPEDDNGRYFNFRALKGQLLRSELYGLDDSTNKHVPVTVTEFR 893
Db 841 GFTFRFTLWKEGKDVLPDTPEDDNDYWLNRALKQPLRSELYGLDGSQQQIPITVTESR 900
Qy 894 SQVRLQHTDSRYFVLMSSVVESSRYHYERIASDPQCSQNTITLSSDRFGQPLKQLSVQYP 953
Db 901 PQVRLQDQATVSPVLMASVVESSRYHYERIIISDPQCNQDITLSSDLFGQPLKQVSVQYP 960
Qy 954 RRQOPAINLYPDTLPDKLLANSYDDQQLBLTVQSSWHHLTNTVTVLGLPSTRSDI 1013
Db 961 RNRPTTNPYPDTLPDTLTFASSYDDQQQLRLTLTCRQSSWHHLIGNELRVLGLPDTGTRSDA 1020
Qy 1014 FTYGAENVAPAGLNLELLSDKNSLIADDPKREYLQOQKATYTDGONTTPILOTPTROALIA 1073
Db 1021 FTYDAKVVDGLNLETCAENSLIADDPKREYLNQKRTFYTDGKNQTPKTPTRQALIA 1080
Qy 1074 FTETTVFNQSLTSAFNGSPISDKLSTTLEQAGYQOQTNLYFPRTGDKVVAHHGYTDYGT 1133
Db 1081 FTETAVLTESLSAFDGGITPDELPGILTQAGYQOQEPYLPFRTGKVKVVARQGYTDYGT 1140
Qy 1134 AAQFWRPQKQNTQLTGKITLIWANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1193
Db 1141 EAQFWRPVPAQRNSLLTGKMTLKWDTHTYCVITQTDAAAGLTYSANYDWRFLTPVQLTDIND 1200
Qy 1194 NOHLITLDALGRPTTLRFWGTENGWMTGYSSPEKASPPSPDVNAATELKXPLPVAOCOV 1253
Db 1201 NVHLITLDALGRPTVQRFWGLSGVATGYSSSEKFPSPFNDDITAINLTGFLPVAQCLV 1260
Qy 1254 YAPESWMPVLQSKTFNRLAEQDWKLYNARIITEDGRICITLAIYRRWQSQKAIPLQLISLL 1313
Db 1261 YAPDSWMPFLSQETFNLTQBEQETLDRSRIITEDWRICALTRRRWLQSQKISPLVKLL 1320
Qy 1314 NNGRPLPHSLTLTDRYDHPDQIQOQVFSDFGFLQLQAARHEAGMARQNEGSL 1373
Db 1321 TNSIGLPHNLTLTDRYDRDSEQIQOQVAFSDGFGLLQASVRHEAGEAWQRNODGSL 1380
Qy 1374 IINVOHTENWAVTGRTEYDNGKQPIRTYQVFLNDWRYVNSDSARQEKAYADTHVYDP 1433
Db 1381 VTKVENTKRWAVTGRTEYDNGKQPIRTYQVFLNDWRYVNSDSAR--KEAYADTHVYDP 1438
Qy 1434 IGREIKVITAKGWRFTLTFTFWFTVNEDENTAAE 1468
Db 1439 IGREIRVITAKGLRQSQYFPWFVSEDENTAAD 1473

```

```

RESULT 4
US-10-262-794A-32
; Sequence 32, Application US/10262794A
; Publication No. US20030207806A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN

```

```

; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/262,794A
; FILING DATE: 02-OCT-2002
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,567
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-10-262-794A-32

```

```

Query Match 58.0%; Score 4581.5; DB 4; Length 1485;
Best Local Similarity 57.6%; Pred. No. 0;
Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;
Qy 1 MONSDPFSITELSLPKGGGALTGMGEALTPTGPGMAALSPLPISAGRGYAPAFITLNN 60
Db 1 MQDSPFVSITITSLPKGGGALTGMGEALTPTGPGMAALSPLPISAGRGYAPAFITLNN 60
Qy 61 SGAGNSPFLGWDGNCNMTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDE 113
Db 61 NSAGNGPFGIGWQCGVMSISRRTHQGIPOYGNDTFLSPQGEVNMIALNDQGPDIRQDV 120
Qy 114 STLOGINLGATPTVTGYSRLESFHSLEYWQPKT--TGKTDFWLIYSPDQGVHLGKSP 171
Db 121 KTLQGVTLPISTVTRYQARQILDPSKIEYQWQASGOEGRA-FWLITPDGHLHLGKTA 179
Qy 172 QARISNPSTQTQTAQWLLEASVSRGEQIYYQYRAEDDTGCEADBITHHLQATQRYLHI 231
Db 180 QACLANQNDQOIAQWLLEETVTPAGEHVSQYRAEDEAHCCDNKTAHPNVTAQRYLVQ 239
Qy 232 VYGNRTASETLPLGLDGSAPSQADWLFLVLPDYGERSNNLKTTPAFST-TGSWLCQRDPF 290
Db 240 VNYGNIKPQASFLVLDNAPPAPEEWFLFLVFDHGRDTSLSHTVPTWDAGTAQWSVRPDI 299
Qy 291 SRYEYGEIRTRRLCRQVLMVHHLQALDSDSKITEHNGPTLVSRLLIINYDESAIATLVFVR 350
Db 300 SRYEYGEVTRRLRCQQVLMVHHLQALDSDSKITEHNGPTLVSRLLIINYDESAIATLVFVR 359

```







```
QY 651 PEGRLFDHTCOLQWADVGLVSLVSVPHMSPHHWRCDLTNNKFWLLNEMNNMGVHH 710
Db 663 PEGWFDNTCOLQVSDIQLGASIVLVPHWPHWYDFTHNKPWLLNINNRGAET 722
QY 711 TLRVSSQFWLDEKAAATTTGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYARGA 770
Db 723 TLFYSSAQFWLDEKSIQLGKFAASVLPPIHLLWNEALDEITGNRLTKVWNYAHA 782
QY 771 WDEREREFGYVEQTSQLOACNAPERTP-----PALTKWVATGLPVIDNALSTEX 825
Db 783 WDEREREFGYVEQTSQLOACNAPERTP-----PALTKWVATGLPVIDNALSTEX 841
QY 826 WR-DQOAFAGSPRTTW-----QDNKDVPLTPEDNSRYWFRNALKGQLLRSELYGLD 878
Db 842 WRGDDOAFAGTPRTTRVEKNAGOGGDTPIKEPTETEAYWLNAPAMKQLLRSELYG-D 900
QY 879 DSTNK-HVPTVTTEPRSVRRLOHTDSRYFVLWSSVBSRNHYHRIASDPQCSQNTILS 937
Db 901 DKTEKAKIPYVTVEARCVRLIPSNDEAAPSWSITSIIENRSYHVERIIVVDPSCQVVLK 960
QY 938 SDRGQPLKQLSVQYPRQOQAINLYPDTLPDKLLANSYDQOQRLRTTYQSSWHHLTN 997
Db 961 ADEYGFPLAKVDIAYPRNKPAQNPYDLSPLDTLPFADSYDQOQKOLYTKQOQSYHHLTQ 1020
QY 998 NTVRLGLPDSRTSDIFTY-----GAENVPPAGGLNLELLSDKNSLIADDPREYLGQOKTA 1053
Db 1021 QDDWVLGLTDRYSEVHYAQTDAQSDIPKAGLILELLKVDGLIGKDKTYILGQORVA 1080
QY 1054 YTDGQNTPLQTPTRQALIAFTTTFVFNOSTLSAFNGSIPSDKLTTLLEAGYQOQNTYLF 1113
Db 1081 YVGGD-----AEKPTRQVAVVETATAAFDDNALHAFDGVIAPELTLQQLLAGY----LLV 1132
QY 1114 PRTE-----DKWVAHGYTDYGTAAQFPWRQKQSNQTLTKITLIWANDYCVVQTRD 1168
Db 1133 PQISDVAGSSEKVVWVARGQYTEYGSAAQFYRPLIQKSLTKGKTYLSWDTHYCVVVKTED 1192
QY 1169 AAGLTSAKYDWRFLTPVQLTIDNDNHLITLDALGRPITLRFWGTENGKMTGYSSEPKA 1228
Db 1193 GAGMTQAKYDYRFLPQALTDINDNQHIVFNALGQVTSRFRWGTENGKISGYSPEK 1252
QY 1229 SFSPSDVNAAELEKPLPVAQCVYAPESMMPVLSQTFN-----RLABQDWOKLYNARI 1285
Db 1253 PFTVPDTEKALALQPTIPVSCNIYVPSWMLLPQOQLTGQKEGETLWNLHRAQV 1312
QY 1286 TEDGRICTAYRRWVQSKAIPQ-----LISLNGPRLPPHSLTLTTDRYDHDPEQOIRQ 1341
Db 1313 TEDGLICELAYRRWIKRQATSSMAVTLQOILAQTPRQPPHAMTITTDYDSDSQOQLRQ 1372
QY 1342 QVWPSDGFRLLOAAAREHAGMARONEDGSLIIN-----VOHTENRWAVTGRTEYDNK 1395
Db 1373 SIVLSDFGRVLOSQRHEAGWQRAEDGSLIVDNTGKPVVANTTTWAVSGRTEYDGK 1432
QY 1396 QPIRTYQYFINDRWYVNSDARQEKAYADTHYDPIGRBKIVITAKGFRFTLFTPW 1455
Db 1433 QGAIKAYLPYLNDRWYVNSDARSARD--DLYADTHFYDPLGREYQVKTAKGFRWENMFMP 1490
QY 1456 FTVNEDENDTAAEV 1469
Db 1491 FTVNEDENDTAAEL 1504
```

## RESULT 7

```
US-10-754-115-60
; Sequence 60, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
```

```
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 60
; LENGTH: 1428
; TYPE: PRT
; ORGANISM: Serratia entomophila
; US-10-754-115-60
```

```
Query Match 50.3%; Score 3971; DB 4; Length 1428;
Best Local Similarity 52.7%; Pred. No. 2.4e-304;
Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;
```

```
QY 1 MONSDPSITSLSPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGYAPATILNVN 60
Db 1 MONHQDWAITAPTLPSSGGAVTGLKGDIAAAGPDGAATLSPLPVPSPGRGYAPTALNVH 60
QY 61 SGAGNSPGLGWCNVMVIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDES 114
Db 61 SRSGNGPPIGWGIGGAQVRRTRNGAPTDTDTFGPDGEVLVPAUTAAGTQBARQNT 120
QY 115 TLOGINIGATFTVGYRSRLESFHSRLEYWOPKTTGKTDFWLIYSPDQVHLLGKSPQAR 174
Db 121 SLGLNPGGSFNVQYRSRTEGSLSRLEWLPADETETEFWVLYTTPDGOVALLGRNAQA 180
QY 175 ISNPSQTTQTAQMLEASVSRGBOIYYQYRAEDDTGCEADBITHHLQATQRYLHIVY 234
Db 181 ISNPTAPTQTAVLMWESSVLTGEMYYQYRAEDDDGCEAERDAHPQAGQRYPVAVY 240
QY 235 GNRITASETLGLDGSASQADWLVFVFDYDERSNNLTPPAFTTGS--WLCRODRPSR 292
Db 241 GNRQAARTLPAL--VSTPSMDSWLFILVFDYDERSVLSSEAPAWQTPGSGEWLCRODPSG 299
QY 293 YEYGFETRLRCQVLMYHHLQALDSKITENHGETLVSRILYLYNDESAIASTLVFVRV 352
Db 300 YEFGNLTRLRCQVLMFYHVLGVLGSSGANDAPALISRLLLDYRESFSLLENVHVQV 359
QY 353 GHEQGNVVTLPPLLEYQDFSPRHHAWQPMQVLANFNATQWQVLDLKGEGLPGLLYQ 412
Db 360 AYSDGTSICALPALALGWQTFPTPTLSAWQTRDDMGKLSLLQYQVLDLNGEGVGVILYQ 419
QY 413 DKGAWYRSARLGEIGSDAVTWKMQPLSVIPLSQSNASLVLDINGDQDWDVITGPGLR 472
Db 420 DSGAWYREYPRVQGGDDPDVATWGAAALPTPALHNSGILADLNGDGRLEWVATGVA 479
QY 473 GYHSORPDGSTRFTPLNALPEVYTHPRAQLADLMAGLSDLVLIGPKSVRLYANTRDGF 532
Db 480 GMYDTEGCDWLHFTPLSALPEYVAHPKAVLADILGAGLTDWVLIGPRSVRLYSKNDGW 539
QY 533 AKGDVVSQSDITLIPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGQP 592
Db 540 NKGETVQOTERLTLPVPGVDPRTLVAFSDMAGSQOHLTEVRANGVRYWPNLGHGRFGQP 599
QY 593 ITLPGSQPATEFNPQAVVYLDLQSGGPTDLIYVHTNRLDILFNKSGNGFPEPVTLRPE 652
Db 600 VNIPIGFSQSVTTFNPDQILLADTDGSGTDLIYAMSRLVIFYFNOSGNYPFEPHTLLPK 659
QY 653 GLRFDHTCOLQWADVQGLGVASLILSVPHMSPHHWRCDLTNNKFWLLNEMNNMGVHHTL 712
Db 660 GVRDRTCSLQVADIQGLGVESLTLTVEHVAHPHWHVCHLSADKPWLLNGMNNMGARHAL 719
QY 713 RYRSSQFWLDEKAAALTTGQTPVCYLPPIHTLMQTEDEISGNKLVTLRYARGAWD 772
```

Db 720 HYRSSQVFLDKAEALAAAGSSPACLYLFTLTLWRSVVQDEITGNRLNSDVLRYHGVD 779  
Qy 773 GREPEFGFVGEQTDHOLA-QGNAPERTPPALTKNWYATGLPIVDNALSTEWYRDD-Q 830  
Db 780 GQERFGRFGVEIRDTDTLASQGTATSELSMPSVRNRYATGVPADVDERUPETTYWQNDAA 839  
Qy 831 AFAGFSPRFTTWQDKOVPLTPEDNSKRYWPNRALKGQLRSELYGLDDSTNKHVPYTVT 890  
Db 840 AFADPATRFTVSGSEDEQTYTP-DDSKTFWLRQALKGILLSELYGADGSSQADIPISVT 898  
Qy 891 EFRSQVRLQHTDSRYVPLVMSVSVESRNYHYERIASDPQCSQNTILSSDRFGQPLKQLSV 950  
Db 899 ESRPQV-RLVEANGDYVPMWMAESRTSVYRVHNDPQCCQQAALLSDEYGFPLRQVSV 957  
Qy 951 QYPRQOPAINLYPDTLPDKLLANSYDDQOLRLTYQQSSWHLLTNTVTR--VLGLPDS 1008  
Db 958 NYPRRPSADNYPASLPATLIFANSYDEQQQLRLGLQSSAHLVLSLSEGHWLLGLAEA 1017  
Qy 1009 TRSDIFTYGARNVPAGGLNLELLSDKNSLIADDPREVLGQOKTAYTDGQNTTLPQTPT 1068  
Db 1018 SRDVFYTSADNVEGGTLLSHLLAPESLSVSDSQVGTLAGQQQWYVLDSDQVATVAAPPL 1077  
Qy 1069 QALIAFTETTVFNQSTLSAFNGSIPSDKLSLTLEAGYQQTNYLFP--RTGEDKVVVAHH 1126  
Db 1078 PPKVAFIETAVLDGWSVSLAAIYVDEH---LEQAGVRSQGYLFPGRGAEQALWTQCQ 1133  
Qy 1127 GYTDYGTAAQWRKQOSNTOLTKITLIWDANCVVVVQTRDAAGLTSAKYDWRFLTPV 1186  
Db 1134 GYVTYAGAHEFWLPSFRSDMLTGPTVTVTRDAYDCVITQWQDAAGIVTTADYDWRFLTPV 1193  
Qy 1187 QLTINDNOHLITLDALGRPTLTFWGTENGKMTGYSSPEKASPPSDVNAALTELKPL 1246  
Db 1194 RVDPNQNLQVLDALGRVTLTFWGTENGIAITYSD---ATLSDVDPGAAALALATPL 1250  
Qy 1247 PVAQCQVYAPESMPLVLSQKTFNRLAEQDMQKLYNARIITEDGRICTLAYRRVQSOKAI 1306  
Db 1251 PVAQCLVVTDSW-----GDDNEK----- 1270  
Qy 1307 PQLISLNLNGRPLPHSLTLTDRYDHPBEOIIRQOVVFSDFGRGLLOAAARHAGMARQ 1366  
Db 1271 -----MPPHVVLATDRYSDTGQOVRQQTFSDFGRGLQSAQTRQAGNAWQ 1318  
Qy 1367 RNEDGSLI-----INVQHTENWAVTGRTEYDNKGOPIRTYOPYFLNDWRVYNSDSAR 1419  
Db 1319 RGRDGLVTASDGLPVTVTA-TNFRWATGRABYDNKGLPVRYQPYFLDSQYVSDDSAR 1377  
Qy 1420 QEKEAYADTHYVDPIGREIKVITAKGWFRRTLFTFPWFVNEDNDT 1465  
Db 1378 Q--DLXADTHFYDPTAREWQVITAKGERRQVLYTFWFVVSDENDT 1421

## RESULT 8

US-10-753-901-18  
; Sequence 18, Application US/10753901  
; Publication No. US20040194164A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Larrinua, Ignacio  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Schafer, Barry  
; APPLICANT: Bevan, Scott  
; APPLICANT: Young, Scott  
; APPLICANT: Guo, Lining  
; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control  
; FILE REFERENCE: DAS-105X  
; CURRENT APPLICATION NUMBER: US/10/753,901  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR APPLICATION NUMBER: US 60/441,717  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 18

; LENGTH: 1493  
; TYPE: PRT  
; ORGANISM: Xenorhabdus nematophilus  
US-10-753-901-18  
Query Match 48.5%; Score 3834; DB 4; Length 1493;  
Best Local Similarity 50.9%; Pred. No. 1.8e-293;  
Matches 767; Conservative 213; Mismatches 472; Indels 56; Gaps 21;  
Qy 1 MQNSQDSITELSLPKGGGAIKMGGEALTPTGPDGMAALSILPLISAGRGVAPAFITLNYN 60  
Db 1 MQGSTPLKLEIPSLPGSGSLKGMGEALNAVGAEGGASFLPLPISVGRGLVPVLSLNY 60  
Qy 61 SGAGNSPGLGDCNVMVIRRRHFGVPHYDETDTLFGPEGEVL-VVAD---QP--RDES 114  
Db 61 STANGSGFMWCGVGFISLRTAKGVPHPTGQDEYLGPDGEVLISIVPDSQCGEQRTAT 120  
Qy 115 TLOGINILGATFTVTGYSRLESFHSRLYEQP--KTTGKTDFWLIYSPDGOVHLLKSPQ 172  
Db 121 SLLGTVLTPHTVTRYQSRVAEKIVRLEHWQPPQORREETSFWLFTADGLVHLFGKHH 180  
Qy 173 ARISNPQTTQTAOWLLEASVSRSRGEQIYYQYRAEDDTGCEADEITHLQATAQRYLHV 232  
Db 181 ARIADPQDETIRIARWLMMEETVTHTEHIIYYHRAEDDLDCDEHQAQHSQVTAQRYLAKV 240  
Qy 233 YYGNRTASETLPGLDGSPADWLFYLVDFDYGRSNNLKTTPAFSTTGS----- 282  
Db 241 SYGNTQETAFFAVKSGIPADNWLFLVDFDYGRSSLSNVPEFNVSENNVPEK 300  
Qy 283 WLCKRQDRFSRYEYGFETRRLCRQVLMYHHLQALDSKITEHNGPTILVSRILNYDSAI 342  
Db 301 WRCKRQDRFSRYEYGFETRRLCRQVLMYHHLQALDSKITEHNGPTILVSRILNYDLNNK 360  
Qy 343 ASTLVFVRVCHQDGNVVTLPLELAYQDFSPRHHAHQWQMDVLAFNAIQRLQVLDLK 402  
Db 361 VSLLTARLAHAETDGTVPVMSPLEMDYQRVNHGVNLNWQSMQPLEKONTLPQVLDLY 420  
Qy 403 GEGVLGLLYQD-KGAWYRSAQR-LGEIGSDAVTWKMQPLSVIPSQSNASLVDINDGDG 460  
Db 421 GEGISGLYQDTQAWWYRAPRVDITAEGTNAVYEEAKPLPHIPAQOESNMLDINDGDG 480  
Qy 461 QLDWVITGPGLRGYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLADLADLADLADLADL 520  
Db 481 RLDWVITASGLRGVHTWSPEGEWTFPIPLSAVPMYEPHPQAKLADIDGAGLPDALIGPN 540  
Qy 521 SVRLYATRGDFAKGVQVQSGDITLVPFGADPKLVAFSDVLGSGQAHLEVSATKVT 580  
Db 541 SVRVSNRAGWDRAQDVIIHLSMDPLEVPGRNERHLVAFSDMTSGSGSHLVEVTADSVRY 600  
Qy 581 WPNLGRGRFGOPITLPGFSQAPATEFNPVPAQVYADLDGSGPTDLIYVHTNRLDIPLNKS 640  
Db 601 WPNLGRGRFGOPITLPGFSQAPATEFNPVPAQVYADLDGSGPTDLIYVHTNRLDIPLNKS 660  
Qy 641 GFAPVTLRFPPEGLRFDHTCOLQADYVGLGVASLILSVPHMSPHHWRCDLTNNKPMLLN 700  
Db 661 HFAEPQRIIDLPGVRFDDTCLRLQIADTQGLGTASIIITLPHMKVQHWRLDMTIFKPMLLN 720  
Qy 701 EMNNMGVHHTLYRSSSQFWLDEKAAALTGTQTPVCLPPPIHTLWQTEDEISGNKL 760  
Db 721 AVNNMGTTETLYYRSSAQFWLDEKQASESGMTWVSYLPPFPVHLWRTVDELDISGNRL 780  
Qy 761 VTTLRYARGAWDGRERFRGFGVVEQTDHOLA---QGNAPERTPPALTKNWYATGLPVI 817  
Db 781 TSHVHYSHGAWDGLERFRGFRVGTQDIDSRASATQGTAEPPAPSPRTVWYGTGVREV 840  
Qy 818 DNALSTYWR-DDQAFAGSPRFTTWQDNK--DVLPTEDDNSRYFNRAKQLLRSEL 874  
Db 841 DILLPTEYWGQDQAFPHFTPRFTRYDEKSGDGMTVP-SEQEYWLHRAKQLRSEL 899  
Qy 875 YGLDDSTNKHVPYTVTFEFSQVRLOHTDSRYPVLMSVSVESRNYHYERIASDPQCSQNI 934  
Db 900 YGDDDSILAGTPYSVDSERTQVRLLPVMSVDPVAVLSVAESRQYRYERVATDPCQSKI 959

QY 935 TLSSDRFGQPLKQLSVQVPRQOPAINLYPDTLPDKLLANSYDQOQRLRLTYQSSWHH 994  
 DB 960 VLKSDALGFPQDNLEIAYSRPQEPFSPYPTLPETLFTSSDFEQMFLRLTRQSSYHH 1019  
 QY 995 LT-NNTVRVLGLPSTSRSDIPTYGAENVPAGGLNLELLS--DKNSLIADDPREVILGOOK 1051  
 DB 1020 LNHDNTWITGLMTSRSDARIYQADKVPDGGFSLWFESATGAGALLLPDAAADYLGHQR 1079  
 QY 1052 TAYTDGQNTTLPQPTTQALIAFTETTVFNOSTLSAFNSIPSKLSLTTLBOAQVQOYNY 1111  
 DB 1080 VAYTG----FEEQPAIPPLVAYIETAEFSDERSLAAPBEVMDQBELTKQLANDAGNNTAKV 1134  
 QY 1112 LFPRTGDKVVAHGYTDYGTAAQFWRPQOSNTQLTKITLWANDYCVVQTRDAAG 1171  
 DB 1135 PFEKTDPHVWVGQKEFTEYAGADGFYRPLVQRETGLTKGTVTWDSHYCVITATEDAAG 1194  
 QY 1172 LTTSAKYDMRFLTPVQLTDINDNHLITLDALGRPILRFGWTENGKMTGYSSPEKAS-- 1229  
 DB 1195 LRMQAHYDIREFWADNTTVDNDNHTVTFDLAGRVTSRFRWGTENGKQGTTPAENETVP 1254  
 QY 1230 FSPSDVNAALIELKKPLPVAOCQVYAPESMWPVLSQKTFNRLAEQDMOKLYNARIITBDG 1289  
 DB 1255 FIVPTVDDALAKPGIPVAGLWVYAPLSWM---VQASFSNDGEL-YGELKXPAGITEDG 1310  
 QY 1290 RICTLAYRWVQSO--KAIPOLISLLNNGPRLPHSLTLTTRDYDHDPEQOIRQOVVPSD 1347  
 DB 1311 YLLSLAFRWQNNPAAAMPKQVNSQ-----PPHVLVSIVITDRYDADPEOQLRQTFTFSD 1365  
 QY 1348 GFGRLQAAARHAGMARQEDGSLII-----NVQHTENRWAVTGTEYDNDKGOPIRT 1401  
 DB 1366 GFGRLQAVRHESGEAWRDEYGAIVAEHGAETAMTDFRNVAHSVTEYDGGQALRK 1425  
 QY 1402 YQPYFLNDWRVYNSDSRQKEAVADTHVYDPIGREIKVITAKGWFRTLTFTPMFTVNEED 1461  
 DB 1426 YQPYFLNSQVYSDSRAQ--DIYADTHYDPLGREYQVITAKGFGRESLTFTPMFVNVNEED 1483  
 QY 1462 ENDTAAEV 1469  
 DB 1484 ENDTAGEM 1491

## RESULT 9

US-10-754-115-18  
 ; Sequence 18, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintrim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Ni, Weiting  
 ; APPLICANT: Zhu, Baolong  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Meade, Thomas  
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
 ; FILE REFERENCE: DAS-104XCI  
 ; CURRENT APPLICATION NUMBER: US/10/754,115  
 ; CURRENT FILING DATE: 2004-01-07  
 ; PRIOR APPLICATION NUMBER: US 60/441,723  
 ; PRIOR FILING DATE: 2003-01-21  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 18  
 ; LENGTH: 1493  
 ; TYPE: PRT  
 ; ORGANISM: Xenorhabdus nematophilus  
 US-10-754-115-18

Query Match 48.5%; Score 3834; DB 4; Length 1493;  
 Best Local Similarity 50.9%; Pred. No. 1.8e-293;

Matches 767; Conservative 213; Mismatches 472; Indels 56; Gaps 21;  
 QY 1 MNSQDSFIITELSLPKGGGALTGMGEALTPTGPGDMAALSPLPISAGRGVAPAFITLYN 60  
 DB 1 MQGSTPLKLEIPLSPSGGSLKMGGEALNAVAGAGGASFSLPLFISVGRGLVPVLSLNS 60  
 QY 61 SGAGNSPGLGWCNVMITIRRTTHFGVPHYDETITELGPEGEVL-VVAD---QP--RDES 114  
 DB 61 STAGNGSGMGWQCGVGFISLRTAKGVPHYTGQDEYLGDEGVLSIVPDSQGPQRTAT 120  
 QY 115 TLOGINIGATFTVGYRSRLESFHSRLEYWQP--KTTGKTDFWLIYSPDGOVHLGLKSPQ 172  
 DB 121 SLLGTVLTPQHTVTRYQSRVAEKIVRLHEWQPQRRBEETSFWLFTADGLVHLFGKHH 180  
 QY 173 ARINSPQTTQTAQWLLEASVSRSQEIYYOYRAEDDGTGCEADEITHLQATQAQRYLHV 232  
 DB 181 ARIADPDQETRIARWLMEETVTHTGEHYHYRAEDDLDCEHELAHQSVTAQRYLAKV 240  
 QY 233 YNGRTASETLPLGLDGSAPSQADMLFVLVDPYDGRSNNLKTTPPAFTTGS----- 282  
 DB 241 SYGNTQPETAFPAVKSGIPADNDWLFLVDPYDGRSSLSNVPEFNVSENNVSENNVPEK 300  
 QY 283 WLCRQDSFRSRYEYGFETRRLCRQVLMYHHLQALDSKITEHNGPTLVSRILINYDESAI 342  
 DB 301 WRCPDPSFRSRYEYGFETRRLCRQVLMFHLQALAGEKVAEETPALVSRILIDYLNKK 360  
 QY 343 ASTLVFVRVGHGEOQGNVVTLPPELAYQDPSPRHHAHQMDVLVANFNAQRMQLVDLK 402  
 DB 361 VSLLTQARRLAHETDGTVMMSPLEMDYQRVNHHGVNLMQSMQPLEKNTLQPYQLVDLY 420  
 QY 403 GEGPLGLLYOD-KGAWYRSRQ--LGEIGSDAVTWKMQPLSVIPSLQSNASLVDINGDG 460  
 DB 421 GEGISGVLYQDTQKAWYRADVDITAGTNAVITYEAKPLPHIPAQGESAMLLDINGDG 480  
 QY 461 QLDWVITGPGURGVHSORPGSWTRFTPLNALPVEYTHPRQAQADLMGAGSLDLIGPK 520  
 DB 481 RLDWVITAGLGRVHTMSPEGEWTFPLSAVPMYFHPQAKLADIDGALPDALIGPN 540  
 QY 521 SVRLYANTRDFAKGKDVQSGDITLVPVGADPKLVAFSDVLSGQAHLEVSATKVTVC 580  
 DB 541 SVRVSNRNGWDRADQVHLSDMPLPVPGRNERHLVAFSDMTGSGQSHLVEVTADSVRY 600  
 QY 581 WPNLGRGFGQPIYLPGESQAPATEFNPQAQVYADLDGSGPTDLIVVHTNRDLIDFLNKG 640  
 DB 601 WPNLGHGKFGEPMLMTGFISETFNPDRLYMVDIDGSGTDFIYARNTYLEYANESGN 660  
 QY 641 GFASPVTLLRPRGLRFDHTCOLQADVQGLGVASILLISVPHMSPHHRCDLTNMKPWLIN 700  
 DB 661 HFAEPQRIIDLPDGVRFDDTCRLQIADTQGLGTASILLITPHMKVQHWRDLMTIFPWLIN 720  
 QY 701 EMNNMGVHHTLRYSSSQFWLDEKAAALTTGQTPVCYLPPIHTLWQTEDEISGNKL 760  
 DB 721 AVNNMGTTETLYYRSSAQFWLDEKLAQESGWTVSVYLPFPVHVLRTEVLDEISGNRL 780  
 QY 761 VTLRYARGADGRERPRGRGYEQTDSHOLA---QGNAPERTPALTKWYATGLPVI 817  
 DB 781 TSHYHSHGAWDGLEREPGRGRVVTQTDIDSRASATQGTAEPPAPSPVTVMYGVGVREV 840  
 QY 818 DNALSTYWR--DQAFAGFSFRFTTWQDNK--DVLPTEDDNSRYWFNRALKGLRLSEL 874  
 DB 841 DILLPTETWQDQQAFFHFTFRTRYDEKSGDMVTVP-SEQEYFWLHRLKGLRLSEL 899  
 QY 875 YGLDDSTNKHVPYTVTFERSQVRLOHTDSRYPLVMSVSVSRNHYHRIASDPQCSQNI 934  
 DB 900 YGDDDSILAGTFYSVDESRTQVRLLPVMVSDVPAVLVSAESRQYRERVATDPQCSQKI 959  
 QY 935 TLSSDRFGQPLKQLSVQVPRQOPAINLYPDTLPDKLLANSYDQOQRLRLTYQSSWHH 994  
 DB 960 VLKSDALGFPQDNLEIAYSRPQEPFSPYPTLPETLFTSSDFEQMFLRLTRQSSYHH 1019  
 QY 995 LT-NNTVRVLGLPSTSRSDIPTYGAENVPAGGLNLELLS--DKNSLIADDPREVILGOOK 1051  
 DB 1020 LNHDNTWITGLMTSRSDARIYQADKVPDGGFSLWFESATGAGALLLPDAAADYLGHQR 1079

```
Qy 1052 TAYTDGQNTTPLQTPTRQALIAFTTETVPNSTLSAENGSIIPSKLSTTLLEQAGYQQTNY 1111
|||
Db 1080 VAYTG-----PEEQPAIPPLVAYIETAEFDRSLAAFEENVDEQELTKQLNDAGWNATKV 1134
|||
Qy 1112 LFPRTGDKVVAHHGYTDYGTAAQFWRPQKQSNLTQTKITLIWDANVCVVVQTRDAAG 1171
|||
Db 1135 PFSEKTDPHVWVGQKEFTEYAGAGFYRPLVQRETGLTKITVTWDSHYCVITATEDAAG 1194
|||
Qy 1172 LTTSKADYDWRFLTPVQITLDINDNOHLITLDALGRPITLRFWGTENGKWTGYSSEKAS-- 1229
|||
Db 1195 LRMOAHYDYRPMWADNTDNDVNDYHTVTFDALGRVTSFRFWGTENGKQGYTPAENETVP 1254
|||
Qy 1230 FSPSDVNAATELKKPLPVACQVYAPESWMPVLSOKTENLAEODWQKLYNARIITEDG 1289
|||
Db 1255 FIVPTTDDALALRPGIPVAGLWYAPLSMW---VQASFSNDGEL-YGELKPAGITIEDG 1310
|||
Qy 1290 RICTLAVRRVQSQ--KAIPQLISLANNPRLPSPHSITLTDRYDHPDEQIQOQVVFSD 1347
|||
Db 1311 YLLSLAFRRQNNPAAAMPKQVNSQN-----PHVLSVITDRYDADPEQQLRQTFTFSD 1365
|||
Qy 1348 GFGHLLQAAHEAGMARQNEGSLII-----NVOHTENRWAIVTGRTEYDNKGQPIRT 1401
|||
Db 1366 GFGRSLQTAVRHESGEAWVRDEYGAIVAEHNHGAPEATMTDPRWAVSGRTEYDGRGOALRK 1425
|||
Qy 1402 YQPYFLNDWRVVSNDARSQKEAYADTHVYDPIGREIKVITAKGFRRLTFTPMFTVNEED 1461
|||
Db 1426 YQPYFLNSWQVSDSARQ--DIYADTHYDPLGREYQVITAKGGRFRSLRFTPMFVWVED 1483
|||
Qy 1462 ENDTAAEV 1469
|||
Db 1484 ENDTAGEM 1491
|||

RESULT 10
US-10-609-113-11
; Sequence 11, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; FILE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1444
; TYPE: PRT
; ORGANISM: Paenibacillus strain IDAS 1529
US-10-609-113-11

Query Match 39.3%; Score 3106.5; DB 4; Length 1444;
Best Local Similarity 43.0%; Pred. No. 7.1e-236;
Matches 634; Conservative 230; Mismatches 551; Indels 61; Gaps 17;

Qy 3 NSDPSITELSLPKGGGAIQMGALITPTGPDGMAALSPLPISAGRGYAPFTLNTNSG 62
|||
Db 5 SNADIKLLSPSLPKGGGSMKGIENIAAPGSDGMARCNVLPVTSGRYITPDISLSVASG 64
|||
Qy 63 AGNSPFLGWCNWTMTTERRRTHFGVPHYDETDFTLGEGEVLVVDQPRDE-----STL 116
|||
Db 65 HGNGAYGMGTWGMVNSISRSRGTPTSYTSBDQFLGPDGEVLVLPESNEQGEIITRHTDTA 124
|||
Qy 117 QGINLGATFTVTGYRSRLSHFSRLEVWQPKTTGKT-DFWLIYSPDGQVHLLGKSPQARI 175
|||
```

```
Db 125 QGIPLGTEFTVTRYPRIESAFLHLEYWQAQAGSATASFWLIHSDAGVLHCLGKTAQARI 184
|||
Qy 176 SNPSQTTQTAQWLLLEASVSRGEOIYYQYRAEDDTGCEADEITHLQATAQRIHLIHYVG 235
|||
Db 185 AAPDSDAKIAEWLVESVSPFGEHIYYQYKEEDNQGVNLEEDNH--QYGANRYLKSIRYG 242
|||
Qy 236 NRTASETLPLGLDGSAPSQADWLFLVDFDYGERSNNLKTTPAFSTTGSWLCRQDRFSRYEY 295
|||
Db 243 NKVASPSLYVWKGEIPADGGWLKSVILDYGENDTSDVPPLYTPQGEWLVRPDRFSRYDY 302
|||
Qy 296 GFEIRTRLCRQVLMYHHLQALDSKITEHNGPTLAVSRLLIILNYDESAIASTLVFVRRVGE 355
|||
Db 303 GFEVTRCLCRQVLMFHFVKELGGE-----PALVWRMQLEYDENPAASMLSAVRQLAYE 356
|||
Qy 356 QDGNVVTLPLELAYQDFSPRHHAHWQPMDFLANFNALQRWQLVDLKGEGPLGLLYQDKG 415
|||
Db 357 ADGAIRSLPPELEFDYTFEGTETADWQFLLFVPEWAEDEHYQLVDLYGEGIPGLLYQNND 416
|||
Qy 416 AWYRSARQLGEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDWITGPGLRGYH 475
|||
Db 417 HWHYRSPAR-GDT-PDGIATNSWRPLPHI PWSRNGMLMDLNGDCYLEWLLAEFGVAGRY 474
|||
Qy 476 SORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGPAKG 535
|||
Db 475 SMNPKDMSGFPVPLQALPTFEFFHPQAQLANVTGSLGTLDMVIGPKSVRFVAGEBAGFKRA 534
|||
Qy 536 KDVVQSGDITLVPDGPAPKLVAFSDVLGSGOAHLEVSATKVCWCNHLGREGFGQITL 595
|||
Db 535 CEVMQVQGITLPERVDKKELVAFSDMLGSGQSHLVRHSDGVTWCNHLNGVFGAPLAL 594
|||
Qy 596 PGFSQPATEFNPAQVILADLDGSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRFPEGLR 655
|||
Db 595 HGFTASEREFNPERVILVDLGGSGADIIYASRDALLIYRLNSGNGFADPVRFPLPGVR 654
|||
Qy 656 FDHTCOLQADVQGLGVASLILSVPHMSPHHRCDLTNMKFWLNLNEMNNMNGVHHTLRYR 715
|||
Db 655 FDNLCRLLPADIRGLGVASLVLVHVPYAPRSWKLDPFAAKPYLLQTVSNLNGASSSEFWYR 714
|||
Qy 716 SSOQFWLDEKAAALTTGQTPVCYLPFPIHILWQTEDEISGNKLVTTLRYARAWDRE 775
|||
Db 715 SSTQYWLDEKQAA-----SSAVSALPFPINVSDMHTVDEISGRTRQKTYRHHGVYDRT 770
|||
Qy 776 REPRFGYVQTSQSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWRDDQAFAG 835
|||
Db 771 KEFAGFRIDTWEERESEGTLSVSTPEVLTTRVYHTGQDERAQQVWQGPAAFOV 830
|||
Qy 836 SP-RFTTWQ--DNKDVLTPTEDDNSRYWFNRAKQGLLRSELYGLDSDTNKHVYVTFEF 892
|||
Db 831 KPVELTRFDAAAQDLPLDSNNGQEQYWLRSQGMPLRTEIPAGD--VGGSPPYQVESF 888
|||
Qy 893 RSQVRLQHTDSRYPLWSSVSRNHYERIASDPQCSQNTILSSRFGOPLKQLSVQY 952
|||
Db 889 RYQVRLVQSIDSECVL-PMQLEQLTYNYEQIASDPQCSQIQOQWFEYGVAAQSVTIQY 947
|||
Qy 953 PRQOPAINLYPDTLPDKLLANSYDDQOROLRTYQQSSWHHLNTNTVRVLGPDSTRSD 1012
|||
Db 948 PRAQPEDNPYPTLPDTSWSSSVDSQMLLRLTRQKAYHLADPGWRNLNIHQITLFD 1007
|||
Qy 1013 IFITYGAENVPPAGLNLELLSDKNSLIADDKPREYLQGOQTAITDGTQNTPTLQTTQAL 1072
|||
Db 1008 AFIYSADSVPAEGISAELELLEVDGTLRSSALEAQYGGQSEIYAGGGE-----PDLRALV 1061
|||
Qy 1073 APTETTVFNOSTLSAFNGSIPSKLSTTLBQAGYQQTNYLFPRTGEDKVVVAHHGYTDY 1132
|||
Db 1062 HYTRSAYLDECDLQAYEGVLSDSQNLNSLASSGYSRILGSGDEVDIFVAEGGFTRYA 1121
|||
Qy 1133 TAAQWEPQKQSNLTQTKITLIWDANVCVVVQTRDAAGLTTSKADYDWRFLTPVQLTDIN 1192
|||
Db 1122 DEPNFFRILQOSSLSSGQVLWDDNFCVTSIEDALGNQIQIAYDYRFEALQITDTN 1181
|||
Qy 1193 DNOHLITLDALGRPITLRFWGTENGKWTGYSSEKASFSPPSDVNAATELKKPLPVACQV 1252
|||
```

Db 1182 NNVAQVALDALGRVVYSRTWTGTEGIGTGF-RPE-VFATPETMEQALALASPLFVAVSCC 1239  
 Qy 1253 VYAPESMMPVLVSQKTFNRL---AEQDMOKLYNARIITEDGRICITLAYRRVWVQSOKAIQOL 1309  
 Db 1240 VYDAHSWGTITLAQSELVDPSEKQWFLIDNRLIMPDGRIKSRGDPWLSHLRLLPPAV 1299  
 Qy 1310 ISLNNNGRPLPHSLTLLTDDYDHPDEQIQOQVVFSGPGRLLQAAARHAGMARQRNE 1369  
 Db 1300 GELLSEADRKPPHTVILAADRYDPDPSSQIQASIVFSDGFGRTIQAKRE-----1349  
 Qy 1370 DGSLLINQHTENRWAVTGRTEYDNKGQPIRTYQVFLNDMRYTNSDSARQEAYADTH 1429  
 Db 1350 -----DTRWAIAERVDYDGTGAVIRSFQFYLLDDNNYVGEAV---SSSMYATII 1396  
 Qy 1430 VYDPIGRIBKIVITAKGWFRRTLFTFPWFTVNEDENDT 1465  
 Db 1397 YDALARQLRWNAKGYERRTAFYFPWFTVNEDENDT 1432  
 RESULT 11  
 US-10-754-115-40  
 ; Sequence 40, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintrim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Ni, Weiting  
 ; APPLICANT: Zhu, Baolong  
 ; APPLICANT: Merlo, Don  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Meade, Thomas  
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
 ; FILE REFERENCE: DAS-104XCI  
 ; CURRENT APPLICATION NUMBER: US/10/754,115  
 ; CURRENT FILING DATE: 2004-01-07  
 ; PRIOR APPLICATION NUMBER: US 60/441,723  
 ; PRIOR FILING DATE: 2003-01-21  
 ; NUMBER OF SEQ ID NOS: 64  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 40  
 ; LENGTH: 1444  
 ; TYPE: PRT  
 ; ORGANISM: Paenibacillus strain DAS1529  
 US-10-754-115-40  
 Query Match 39.3%; Score 3106.5; DB 4; Length 1444;  
 Best Local Similarity 43.0%; Pred. No. 7,1e-236;  
 Matches 634; Conservative 230; Mismatches 551; Indels 61; Gaps 17;  
 Qy 3 NSQPSITSLPCKGGGAIATGGEALPTGPDGMAALSPLPIAGRGYAPAFILNYSNG 62  
 Db 5 SNADIKLLSPSLKGGGSMKGEENIAAPGSDGMARCNVPLPVTSGRYITPDISLSYASG 64  
 Qy 63 AGNSPFLGWCNWNVTIRRHFGVPHYDETDFTLGPGEVLVADQPRDE-----STL 116  
 Db 65 HNGAYGMGTWGMVMSISRRTSRGTPSYTSDEQFLGPDGDEVLPESNEQGEIIRHDTDA 124  
 Qy 117 QGINLGAFTVTGYSRLESFHSRLLEYWQPKTTGKT-DFWLIYSPDGOVHLHGKSPQARI 175  
 Db 125 QGIPUGETFTVTRFPRIIESAFHLLYWEAQSATASFWLIHSGADVHLGCLGKTAQARI 184  
 Qy 176 SNPSOTTQAWLLEASYSRGEQIYYQYRAEDDTGCBADITHHLQATQRYLHIVYG 235  
 Db 185 AAPDSAKIAEWLVEESVSPGEHIYYQYKEEDNQVNLEEDNH--QYGANRYLKSIRYG 242  
 Qy 236 NRTASETLPGLDGAPSQADMFLYLVFYDYGERSNNLKTTPAPFSTGSMCLCRQDRFSREY 295  
 Db 243 NKVASPSLYWVKGIADQWLYSLVLDYGENDTSDAVPPLYTTPQGEWLVRPDRFSRYDY 302

Qy 296 GFEIRTRRLCRQVLMYHHQLQALDSKITEHNGPTLVSRLLIYNDESIAISTLVFVRRVGH 355  
 Db 303 GFEVTRCLCRQVLMFHFVKELGGE-----PALVWRMQLEYDENPAASMLSAVRQLAYE 356  
 Qy 356 QDGNVVTLPPLLEAYQDFSPRHHAHQWOPMDVLAFNAIQRWQLVDLKGEGPLGLLYQDKG 415  
 Db 357 ADGAIRSUPPLEFDFTPPGIETTADQWQFLPVPEWADEEHQOLVDLYGEGIPGLLYQNND 416  
 Qy 416 AWYRSARQLGEIGSDAVTWBKMQPLSVIPSLQSNASLVDINGDQLDQWVITPGPLRGVH 475  
 Db 417 HWHYRSRPAR-GDT-PDGIAYNSWRPLPHI PVNSRRNGMLMDLNGDGYELWLLAEPCVAGRY 474  
 Qy 476 SQPDGSGWTRTFPLNALPVEVTHPRAQLADLMAGLSDLVLIGPKSVLYYANTRGDFPAK 535  
 Db 475 SMNPDKSGSVFPLQALPTFHFPAQLANTYVTSGLTDLVMIGPKSVFVYAGEEAGFKRA 534  
 Qy 536 KDVVQSGDITLPLVFGADPRKLVAFSDVLGSGOAHLEVSATKVTCPNLRGRFRFOPTIL 595  
 Db 535 CEVWQVIGITLPLVERVDKELVAFSDMLGSGQSHLVRIRHGVTCWPNLGNVGFAPLAL 594  
 Qy 596 PGFSQAPATEFNPAQVYLADLDGSGPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFP 655  
 Db 595 HGFTASEREFNPERVYLVDLGGSGASDIIYASRDALLIYRNLSGNGFADPVVRPLPDGVR 654  
 Qy 656 PDHTCOLQMAADVQGLGVASLILSVPHMSPHWRCDLTNMKPWLLNEMNNMVGHHVTLRYR 715  
 Db 655 PDNLCLLIPADIRGLGVASLVLPYMAPRPSWKLDFFAAKPYLLQTVSNLNGASSFWFR 714  
 Qy 716 SSSQFWLDEKAAALTGTGTPCYLPPFPIHTLMQTEDEDEISGNKLVTLRYARGAWDGR 775  
 Db 715 SSTQYWLDEKQAA-----SSAVSALPFPINNVSDMHTVDEISGRTRTKYTYRHGYDRT 770  
 Qy 776 REFRGFGYVEQTDGSHQAQGNAPERTPALTKNRYATGLPVIDNALSTEYWRDDQAFAG 835  
 Db 771 KEFAGFGRIDTWEEERDESEGLTSVSTPPVLTTRTWYHTGQKQDEERAVQOYWGDDPAAQV 830  
 Qy 836 SP-RFTTWQ--DNKDVPLTPBEDDNRSYWFRNALKGQLRLSELYGLDDSTNKHVPVTVTF 892  
 Db 831 KPVELTRFDAAAQDLPLDSNNGQOEWYLRSLQGMPLRTBIFAGD--VGGSPPYQVESF 888  
 Qy 893 RSQVRRLOHTSRYPVLWSSVVSERNHYEIASDPQCSQNTILSSDRFGQPLKLSVQY 952  
 Db 889 RYQVRLVQSIDSCECVALL-PMQLEQLTYNYEIQASDPQCSQIQOMFDEYGVAAQSVTIQY 947  
 Qy 953 PRQOPAINLYPDTLPDKLLANSYDDOQORLRYTQSSMHLTNTNVRVLGLPSTRSD 1012  
 Db 948 PRRAQPEDNPYRPTLPDTSWSSYDSQOMLRLTRQOKAYHLADPEGWRLNIPIHOTRLD 1007  
 Qy 1013 IFITYGAENVAPAGLNLELLSDKNSLIADDKPREYLQOQKTAITDQONTPTPQTRQALI 1072  
 Db 1008 AFYISADSVPAEGISAEELLEVDGTLRSSALEQAYGGQSEIIVAGGE-----PDLRALV 1061  
 Qy 1073 APTTTFVNTQSTLSAFNGSIPSDKLSLTLEQAGVQOQNTNLYFPRTGEDKVVVAHHGYTDY 1132  
 Db 1062 HYTRASVLDDECLQAYEGVLSDSQLNSLSSGYSARSILGSGDEVDIFVAEQGFTRYA 1121  
 Qy 1133 TAAQFWRPQKSNQTLGKITLWDANYCVVVQTRDAAGLTTSKAYDWRFTFPVOLTIN 1192  
 Db 1122 DEPNFRILGQSSLLSGEQLVTWDDNFCVATSTEDALGNQIQIAYDTRFVEAIGITDTN 1181  
 Qy 1193 DNOHLITDALGRPITLRFWGTENGKMTGYGSPKASFPSPSDVNAALIELKKPLPVAQCC 1252  
 Db 1182 NNVAQVALDALGRVVYSRTWTGTEGIGTGF-RPE-VFATPETMEQALALASPLFVAVSCC 1239  
 Qy 1253 VYAPESMMPVLVSQKTFNRL---AEQDMOKLYNARIITEDGRICITLAYRRVWVQSOKAIQOL 1309  
 Db 1240 VYDAHSWGTITLAQSELVDPSEKQWFLIDNRLIMPDGRIKSRGDPWLSHLRLLPPAV 1299  
 Qy 1310 ISLNNNGRPLPHSLTLLTDDYDHPDEQIQOQVVFSGPGRLLQAAARHAGMARQRNE 1369  
 Db 1300 GELLSEADRKPPHTVILAADRYDPDPSSQIQASIVFSDGFGRTIQAKRE-----1349  
 Qy 1370 DGSLLINQHTENRWAVTGRTEYDNKGQPIRTYQVFLNDMRYTNSDSARQEAYADTH 1429



Db 1350 -----DTRWTAERVDYDGTGAVIRSFQPFYLDWNNYGEBAV--SSSMYATY 1396  
Qy 1430 VYDPIGRKIVITAKGWFRTLTFTPWTVNEDENDT 1465  
Db 1397 YDALARQLRMVNAKGYERTAFYFPTVNEDENDT 1432

## RESULT 12

US-10-609-113-39  
; Sequence 39, Application US/10609113  
; Publication No. US20040110184A1  
; GENERAL INFORMATION:  
; APPLICANT: Bintrim, Scott  
; APPLICANT: Bevan, Scott  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Donald J.  
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from  
; TITLE OF INVENTION: Paenibacillus Species  
; FILE REFERENCE: DAS-101XC2  
; CURRENT APPLICATION NUMBER: US/10/609,113  
; CURRENT FILING DATE: 2003-06-27  
; PRIOR APPLICATION NUMBER: US 60/392,633  
; PRIOR FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: US 60/441,647  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 697  
; TYPE: PRT  
; ORGANISM: Paenibacillus apairius strain DB482  
US-10-609-113-39

Query Match 16.0%; Score 1262; DB 4; Length 697;  
Best Local Similarity 37.0%; Pred. No. 2.5e-90;  
Matches 270; Conservative 119; Mismatches 291; Indels 50; Gaps 11;

Qy 736 VCYLPIPIHLWQTTETDEISGNKLVTLRYAGAWDGRERFRGFGYVEQTDSSHQAQG 795  
Db 7 VCGVFPFVNVSDITHTVDEISGSARIQYTYRNGVYDRTDKEFAGFGHIDTWEEERDEG 66  
Qy 796 NAPTPTPALTKNVAAGLPLVIDNALSTEWYRDQAFAGFSP-RFTTWQ--DNKDVPLTP 852  
Db 67 TLSISTPPLVTRWYHTGQKQDEERAVQVQWQGDPAFQVQVPLTRFDATAQDVLD 126  
Qy 853 EDDNSRYWFRNALKGOLLRSLEY-----GLDDSTNKHVPYTVTFRSQVRLQHTDSRYP 907  
Db 127 PNREEYWLRSRAGMPLRNEIFAGDVVGLP-----PYQVESLRVQVRLMQSTDSECV 179  
Qy 908 VLMSSVVESSRNYHYERTIASDPQCSQNTLSSDRFGQPLKQLSVQYPRQQAIPNLYPDTL 967  
Db 180 TL-PMQLEQLTYNYEQIASDPQCSQIQQWFDEYGVAAQSITIQYPRRAQPEDNFPHTL 238  
Qy 968 PDKLLANSYDQOQLALTYQOSSWHLTNTNVRVLGIPDSTRSDIFTYGAEVNPAGLN 1027  
Db 239 PDTSSWSSYSDSQMLLRLTRQKAYHALDEGPEWRLNIPHQTRUDSFYSADSVPAGGIS 298  
Qy 1028 LELISDNKSLIADKPREYLGQQQXTAYTDGQNTTFLQTPTRQALIAFTETTVFNQSTLSA 1087  
Db 299 AELLGGGTLSRPALEQAYGQSIIYAGGE-----PDSRALVHYTRSAILDEACLOA 352  
Qy 1088 FNGSIPSDKLTLEQAGYQOTNYLFPRTGDKVVAHHGYTDYGTAAQFWRPQKQNTQ 1147  
Db 353 YEGVLSQSLSNLSASGYQSRARILGSDGDEADIFVAEQGFTRYADEQNFRIILGQSSSL 412  
Qy 1148 LTGKITLIWDANYCVVQVTRDAAGLTTSKAYDWEFLPVLQTDNDNOHLITLDALGRPI 1207  
Db 413 LTGEQVITWDDNFCVATYSIEDALGNQIQIAYDIFRVAIQITDANNVNVQSLDALGRV 472  
Qy 1208 TLRFWGTENGWMTGYSPKASFPSPSDVNAAILKKPLPVAQCVQVYAPESWMPVLSOKT 1267  
Db 473 YSRWTGTEEGTGF-RPE-AEFSPPETMEQALALASPLPVASCCVYDAHSMWGTITLIGQ 530

Qy 1268 FNRL---AEQDWOKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLNNGLPPHSL 1324  
Db 531 LSAIVPDSEKQWSFLIANRLIMPGRIRARGRAPWMLQRLLPAPAVAKLLSEADRKPHTV 590  
Qy 1325 TLTTRYDHDPEQIIRQOVVFSDFGRLLQAAARHEAGMARQORNEGSLINVOHTENRW 1384  
Db 591 VLAADRYPDPSQIQASVVFSDGFGRTIQAKR-----ADTRW 629  
Qy 1385 AVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAEVADTHVVDPIGREIKVITAK 1444  
Db 630 AITERIDYDETGAIRSFQPFYIDDMNYGKEAV--SGSMYATIIYYDALARQURMVNAK 687  
Qy 1445 GWFRRTLFTPT 1454  
Db 688 GYERTAFYP 697

## RESULT 13

US-10-156-761-7751  
; Sequence 7751, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 7751  
; LENGTH: 2386  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-7751

Query Match 2.6%; Score 202; DB 4; Length 2386;  
Best Local Similarity 18.3%; Pred. No. 1.7e-05;  
Matches 344; Conservative 184; Mismatches 647; Indels 706; Gaps 89;

Qy 13 SLPKGGCAITG--MGEALTPTG---PDGMAALSPLPISAGR---GVAPAFITLNYNS-- 61  
Db 318 TLASGSSSDAGDYRASTLSTPTGSWEVSTGSGAFTYVPIQLPKPPMGSAPSLSYNSQS 377  
Qy 62 -----GAGN---SPFGLGWCNVMTIRRR-----THFGVPH-----YDDETDFLQPEGEVL 104  
Db 378 VDGRTSASNQASWAGMGWDLNVGYIERRYNCSEDGLPTIGMCDWSPNSAKEPSGAVY 437  
Qy 105 VVADQPRDESTLQGINIGATFTVTG---YRSRLESFHSRLEYWQPKTGT-KT-DFWLIYSP 160  
Db 438 VI-----NLNGVTSSELIDQNTGSGAYHLKNDPGWRVQRLFDGYGAGRDGEYVISTQ 489  
Qy 161 DGQVHLHGKSPQAR-----ISN-----PSQTTQTAQWLLLEASVSSRG 197  
Db 490 DGQRYFPGWGRSEKRTGATASVFTPEVVGNDTGTGCHDQFPPECTQAWRWSLDRAVDANE 549  
Qy 198 EQIYYQYFRAEDDTGCEADEITHLQATAQR-----YLHIVYYG----- 235  
Db 550 VETMYFY-----DKEYNHVRSVANSKAREYVSSGVYKEIQIGWSSQIPDGKLP 599  
Qy 236 -----NRTASE-----TLPLDGSAPSQADWLFYLVFDYGERSNL--KT- 273  
Db 600 KVELSHVNRCTIERQVENDPLRDEPATCPTFDKFTSPYDPEVDLMCDGTSADYNCAKGY 659





```
Qy 254 ADWLFLVDFYGBRSNNLKTTPAPSTTTGSMLCRODRESFYEGFEIRTRBLCRQVLMYHH 313
Db 722 A-----LTGNELO--PAY-----SIYHTKDAFK-----HT 744
Qy 314 LOALDSKITBHGPTLVSRLL--ILNYDESA-----IASTLVFVRVG-----HEODGNVVT 362
Db 745 LTVDDVNTN-----SGLGNVLNNEKANVVKADTKDMYKMLSTLDPLNNSDISVDK 797
Qy 363 LPPELAYQDFSP-----RHHAHWQ-----PMDVLANFNAIORWQ 397
Db 798 LDHINHVLNLSAPDDTAIGNVRAMTNNELIRTSTVMNPSLYETTPSNVLKTRVLPNPIK 857
Qy 398 LV-DLKGEGLPGLLYQ-DKGAWVRS-----AORLGEIGSD-----431
Db 858 VIKYKNSGRTRYIYEAPGKWKMQSVKNSITERSSITQNTPEITFDLYNRGTLPTGTYS 917
Qy 432 ---AVTWKMQPLSVIPSLQSNASLVNDINGDQLDWITGFLRGYHSQRPDGSWTRFTP 488
Db 918 IRYATIWDENSEI--VRTEEQSLs-----HNNLELSYVIT--EDLSGN-----KKFVS 962
Qy 489 LNALPVYTHPRAQLADL-MGAGLSDL-----VLGPKSVRLYANTRDGPAGKDVQV 540
Db 963 VIDVPFKIALAKEVASTLTIGKDAASFDKSDQVDNLG-ESVNLQTMNTAN--FTNSEGIIL 1020
Qy 541 SGDTITLP-----VPGADPRKLVPASD-----VLGSGQAHLEVEVSATKVTCPWN 583
Db 1021 EIIVTIPADNIKTNLTALIPDETKYRVVYTTDVRNGVNSNPTDLTKVTAVKYV---1076
Qy 584 LGRGRFGQP-----ITLFGSQPATEFNPQAQVYLAADLG-----SGPTDLIYV 626
Db 1077 -----FDEPLVLTNGSQSFQTMNVRTVPEDAPILTKAH-SQIFTKGLDNTLWSGNKVELET 1130
Qy 627 HTNRLDI-----FLKSGNGPABPVLTRFPPEGLRFDHTCOLQWADVQGLGVASLILSVPHMS 683
Db 1131 EDNRGDLVVKYTBESGNTIQNSLTSKGNKTEYN-----VSVPMQ-1170
Qy 684 PHWRCDLTNNKPWLLNEMNNMGVHTLTVRRSSQFWLDEKAAALTTGOTPVCLPFP--741
Db 1171 -----IDRLN-----RHYKFRVDNQ--LDPTTHYAKGQTKIVNLIVE 1208
Qy 742 -----PIHTLMQT-----ETEDISGNK-----759
Db 1209 VFEGSVIADYKTTDGEVLSPLATVVVNSQIEGTEYATPATIPDRVTFETTTDQGVKKTIS 1268
Qy 760 --LVTTLRYAGAWDGRERFRGRGYVQTDHSHLOAGNAPERPPAL-----805
Db 1269 YHLISTPENOSGTVVGKQTIIEVHYVY-EPITTYEQIPNDAPQETPVALETVRYVDSGNE 1327
Qy 806 -----TKNWATGLPVIDNALSTE-YWRDQAPAGFSPFT-----840
Db 1328 VQTEEGTHDAPGIIADKQWYTGQTAENGITTHVYQRIQSEIENAPQETPVALEVTYCY 1387
Qy 841 -----TWQDNKQVP-----LTPEDDNSRYWFNRALKGQLLRSELYGLD 878
Db 1388 VDSEGNVEQTEEGTHDAPGIIIGDKWYGTQTTEDGITTHYQK-----IQSE---IP 1438
Qy 879 DSTNKHVPYVTFER-----SQVRLOHTDSRYPLWSSVRSNRYHYERIASDPQCSQNI 935
Db 1439 NEAPQETPVALETVRYVDSGNEVOET-----BEGTHQPPSIIIGDKWYTGQT 1486
Qy 936 LSSDRFGQPLKQLSVQYPRROQPAINLYPDTLPDKLLANSY-DQQRQLRLTYQQSSWHH 994
Db 1487 TTADGI-----TYYVVERIQSEIPNEAPKETPQLETVRYVDGEGNEVOET--BEGTHH 1538
Qy 995 LTNNTVRLGLPDSRSDIFTTYGAENVPAGGLNELLSDKNSLIADDPREYLCQ-QKTA 1053
Db 1539 APG-----IIG-----DKWYTGQTTESGITTHVYERIQSEIPNEAPQETPQLEVTR 1587
Qy 1054 YTDQNTTPTLOTP-----TFQALIAETETTFNQSLTSAFNGSIPSDKL 1097
Db 1588 YVNSEGNEVOETEGTHQPPGIIIGDKWYTGQTTADGITTYVYVYERIQSEIPNEAPKE-T 1646
Qy 1098 STTLEQAGYQQTNYLFPRTGTBDKVVVAHHGYTDYGTAAQFWRPQKQSNQTLTGKITLIWD 1157
```

```
Db 1647 PVQLEVTRYVDTD-----GNEVQETEEGTHQPPGIIGDKW-----QVTGRVT-----1688
Qy 1158 ANYCVVVQTRDAAGLTTSAKYDWRFL-----TPVOL-----TDINDNQHLITLD 1201
Db 1689 -----EKDGITT---YVYERIQSAIPNEAPQETPQLETVRYVDITGNEVQETEE 1735
Qy 1202 ALGRPITL-----RFWG-----TENGMKTGYSPKASPPSDVNAAIELKKPLPVAQCOV 1253
Db 1736 GTHQPRYIIGDKWRYSGVTVTENG-ITKHVY-ERIQSKVPNDA-----POETPVQLEVTRY 1789
Qy 1254 YAPESMMPVLISQKTFNRLAE-----ODWQKLYNARIITEDGRICTLAYRRWV 1300
Db 1790 VDPEG-----NEIQETTEGKHQPPGIIGDRWQ--YTGKVKTEKDQ-IITVYVER-I 1835
Qy 1301 QSKAIPOLLISLNNGRPLPHSLTLTTRDYHDHPDPEQIIRQOVVFSQDFGELLQAAARHE 1360
Db 1836 QSE--IP-----NNPQETPVELEVR-----YVDBGEGNEVOET---1867
Qy 1361 AGMARQORNEGSLIINQVHTENRWAVTGR-TEVDNKGQPIRTYQPYFLNDWRVYVNSDSAR 1419
Db 1868 ---TEGHKQPPSII-----GDRWYTGKVKTEKDQ-----ITTY--VYERIQSKVPNDAPR 1912
Qy 1420 QEKEAYADTHVDPFGIEIKVITAKGWFRRRTLFTPWFTVNEDENDTAAEVKK 1471
Db 1913 VDIDELKITIYVDNNGREI-VPSRKGQLPPEQF-----IGDMQYTGHKIEK 1958

RESULT 15
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 2.3%; Score 181.5; DB 4; Length 2364;
Best Local Similarity 16.8%; Pred. No. 0.00072;
Matches 274; Conservative 180; Mismatches 460; Indels 715; Gaps 77;

Qy 7 FSITELSLPKGGGAIQMGREALTPGP-----DGMALSLPLPISA-GRGYAPAFITNY 59
Db 305 YAVTSSSSSDAGDY---RASTLSPTSGSWEVSTGSGSFNYSVPIQVPAPPMGSAFSLAMS 361
Qy 60 NS-----GAGN---SPFGLGDCNVTITRR---THGCVPH-----YDETDTFLGPE 100
Db 362 DSQSDGRTSASNNQASWSGMDLVNGFIERRYRNCTDGLPTIGMCDSPNSAKEPD 421
Qy 101 GEVLVWADQPRDESTLOGLNLTGATFTVTGYSRSLRSHFSRLEYWQPKT-----TGKT-D 153
Db 422 GAAYVISLNGTTSIELIQDNGSGAY-----HLKNDPGWRVQLRFDGHGGRNGE 470
Qy 154 FWLIYSPDGPVHLLGKSPQARISN-----PSQTTQTAOWILLE 190
```

Db 471 YWISTQGRYFQWGRSERTSTATASVFTVPVVGNDAGBCHDQFPPEPCTQAWRNLD 530  
Qy 191 ASVSRGQIYYQYRAEDDTGCEADEITHLQATQRYLHVYVGNRTASETLPGLDGA 250  
Db 531 RAVDANEVETWYFY-----DKEYNHYSVA----- 555  
Qy 251 PSQADWLFLVFDYGERSNLKTTPAFSTTGSWLCRQDRFSRYEYGFIRTRLCRQVLM 310  
Db 556 -----NTDKAREYSSGVYKE-----IQYGM----- 576  
Qy 311 YHHLQALDSKITEHNGPTLVSRLLINYNDESALASTLVFRRVGHQDGNVVTLPLELAY 370  
Db 577 -----SSQITDGKPAKVELSHVN-----RCIERVOENDPLRDEPAACP 616  
Qy 371 QDPSRHHQPMQMDVLAFNAIQRMQLVDLKEGLPGLLYQDKGAWYRSQAQLGEIGS 430  
Db 617 FDDKPGSYDP-VPVDLMCDGTSAD-----YNACAGKTYF----- 648  
Qy 431 DAVTWKMQPLSVIPSLQSNASLVINDINGDGLDWDVITGFLRGYHSQRPDGSRWTRPTPLN 490  
Db 649 -----PTFFSTMDLMDI----- 660  
Qy 491 ALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFPAKGDVQVQSGDITLPPVG 550  
Db 661 -----KTYVSDQDG--TGWDLVQYQYKYGMPN 686  
Qy 551 ADPRKLVAFSDVIGSGOHLVEVSATKVTCPNLGRGFGPITLPGFSPATEFNPQAV 610  
Db 687 PD-----GTIGKTLWL-----DYIQRT 704  
Qy 611 YLADLDGSGPTDLI--YVHTNRLDIFLNKSGNGPAEPVTLRPFEGRLFDHTCQLQMAVQ 668  
Db 705 Y-----GDG-DDIVLVPVFNFRDLD-NKVG-----AELNFR 736  
Qy 669 GLGVASLILSVPHMSPHWRCDLTWKPWLLNEMNNMGVHHTLRYRSSQFPLDEKAAA 728  
Db 737 -----IKEHGDIGATTYSYGANACDIDHLPQ 766  
Qy 729 LTTGQTPVCYLPPFIHTLWQTEDEISGNKLVTLTRYARGAWDGREREPFGYVEQTD 788  
Db 767 ASNTQD--CY-----WQKWTP-----GETDSKTCWFKKF-LVTQV- 799  
Qy 789 SHLOAQGNAPERTPALTKWVATGLPVIDNALSTE-----YWRDDQAFAGSPRFTWQ 843  
Db 800 -----QVDPVTATTN--QDGAIVMTTSTYTYEDGAGWHFTNDPLIKDEDESWTDMR 846  
Qy 844 DNKOVPLT-----PEDDNRVMPNRALKGOLLARSELVGLDDSTNKHVPVTVTEPRSQVRL 899  
Db 847 GYQEVQVTTGAGAGOKTKKSLYGLSG-----DRTSK----- 879  
Qy 900 QHTDSRYFVLMSSVSVESRNYHYRIASDPQCSQNTLSSDRFGQPLKQLSVQYPRROQPA 959  
Db 880 -----ADASATKTVTV-DGDG----- 895  
Qy 960 INLYPDT--LPDKLLANSY-DDQORQLRLTYQSSWHHLTNTVTVGLPDPSTRSDIFTY 1016  
Db 896 -NNYTDSDWLSGRILSLRDDTGTSHERTYHK-YWDH---NTAQYDGLPDA-----FVR 946  
Qy 1017 GAENVYPAGLNLLELSDKNLILADDKPREYLGQOKTAVTDGONTT---PLOTPT-ROALI 1072  
Db 947 EKE-----TTNTKVVSSCW-REHTVETEDDTGEGASTTFGLPMRDTDDWGQSSV 993  
Qy 1073 AFTETTVF-----NOSTLSAFNGSI FSKLS-----TTLEQAGYQOTNY----- 1111  
Db 994 SDNRCTTYGRAVNTDNYDSTGAQRWTVLQDOVKYHSGVCSSTADSKQDGYASTLYDNATS 1053  
Qy 1112 -----LFPRTGEDK-VWVAHGYTDYCTAAQFWRPQKQSNQTLTG-KITL 1154  
Db 1054 VDANKPVDGNPTESRITYTKSGSYRSTW---SGYDDAGRV---MWSEDSGSHNRTLTKYSPAN 1108  
Qy 1155 IWDANYCVVQTRDAAG-----LTTSAKYDWRFL-TPVOLTINDNQHLITLDALGRPI 1207  
Db 1109 TWPLN-GVITVTPDPDGPALPAHTALTSTAWTSRFGKPTSIQDANGNVTKMSLDAAGRLV 1167

Qy 1208 TLRFW-GTENGKMTGYSSPE-KASFSPSPSDVNAAIELKKPLPVAQCQVYAPESMMPVLSQ 1265  
Db 1168 EV--WRPTETG-----SSPSMKFSYTIPTNSA----- 1194  
Qy 1266 KTFNRLAEQDWQKLYNARIITEDGRICTLAYRRVWQSKAIPQLISLLNNGPRLPPHSI- 1324  
Db 1195 -----GVFDV-----DGYPHVATHVLO 1212  
Qy 1325 TLTTDRYDHDPEQOIROQVWFSDGFGRLLOAAARHEAGM-ARORNEGSLIINVQHTNR 1383  
Db 1213 SGTIVLSSH-----AYVDGLGRARETQTPMGNGVDAATGNEVNRQVSVTRYDSA 1262  
Qy 1384 WAVTGRTE-YDNKG-----QPIRTYQPIFLNDW--RYVSN----- 1415  
Db 1263 GNVGTSAVFRNQGTAGSGSPSSAKVEDLPSYTDLVL-DWAGRAITSLRQVNGASQDAGR 1321  
Qy 1416 -----DSAROEKEAYADTHVVDPIGREIKVITAKGWFRRTLFTPWFTVNEDE--- 1462  
Db 1322 VDTTYDGDFTSVKNKVDAAADTYT-DVYGVQSVKVEHTG--SATYTTAYTTAKDELIKI 1378  
Qy 1463 -----NDTA 1466  
Db 1379 TDPRGNDTS 1387

Search completed: February 16, 2006, 21:54:17  
Job time : 144.704 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 340.476 Seconds  
(without alignments)  
5259.201 Million cell updates/sec

Title: US-10-754-115-34  
Perfect score: 13043  
Sequence: 1 MYSTAVILNKISPTRDQTM.....KALLESLDIILHITYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12396	95.0	2538	Q93RN7	XENNE
2	12098	92.8	2524	Q8KSP0	XENORHABDUS
3	5486.5	42.1	2523	Q93RP0	XENORHABDUS
4	5043	38.7	2516	Q9RN43	PHOLU
5	4986.5	38.2	2504	Q85160	PHOLU
6	4914.5	37.7	2525	Q7N7Y9	PHOLU
7	4802	36.8	2381	Q8GF92	PHOLU
8	4763.5	36.5	2378	Q7N7Y6	PHOLU
9	4583.5	35.1	2376	Q9F923	9ENTR
10	4572	35.1	2499	Q8GFA0	PHOLU
11	4515.5	34.6	2444	Q7N7Y1	PHOLU
12	4491	34.4	2215	Q7N7Y0	PHOLU
13	4061	31.1	2373	Q6XPA4	YERFER
14	3948	30.3	2502	Q4ZNN7	PSESY
15	2965.5	22.7	2168	Q693A5	YEREN
16	1807.5	13.9	1197	Q8D1P7	YERPE
17	1779	13.6	1192	Q6QD23	YERPS
18	1605.5	12.3	1362	Q7N493	PHOLL
19	1595.5	12.2	1564	Q7MZV6	PHOLL
20	1591	12.2	1565	Q85156	PHOLU
21	1571	12.0	1189	Q85152	PHOLU
22	1553.5	11.9	1615	Q87X47	PSESM
23	1507	11.6	1471	Q7N8C3	PHOLL
24	1495.5	11.5	1617	Q4ZP57	PSESY
25	1332	10.2	1279	Q93RP4	XENNE
26	781	6.0	2986	Q4HWU4	GIBZE
27	758	5.8	147	Q8KSP4	XENNE
28	527	4.0	168	Q6WCC2	9ENTR
29	492.5	3.8	139	Q6WCC3	9ENTR
30	440.5	3.4	1095	Q85151	PHOLU
31	410	3.1	1080	Q66PW7	PHOLU

32 355.5 2.7 981 2 Q4ZP58 PSESY  
33 346 2.7 80 2 Q8KSP2 XENNE  
34 342.5 2.5 830 2 Q6QD22 YERPS  
35 317 2.4 833 2 Q8ZAV3 YERPE  
36 309 2.4 774 2 Q8D1P8 YERPE  
37 309 2.4 965 2 Q85155 PHOLU  
38 298 2.3 776 2 Q7APN4 YERPE  
39 292.5 2.2 966 2 Q7MZV5 PHOLL  
40 292 2.2 1173 2 Q7N492 PHOLL  
41 285.5 2.2 779 2 Q8KSP3 XENNE  
42 264.5 2.0 981 2 Q7N8C4 PHOLL  
43 262.5 2.0 10203 2 Q5HPA2 STAEQ  
44 261 2.0 4936 2 Q6D5C2 ERWCT  
45 254.5 2.0 4688 2 Q9PQ08 UREPA

#### ALIGNMENTS

##### RESULT 1

Q93RN7 XENNE PRELIMINARY; PRT; 2538 AA.  
AC Q93RN7  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE XptA2 protein.  
GN Name=XptA2;  
OS Xenorhabdus nematophilus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Xenorhabdus.  
OX NCBI\_TaxID=628;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX MEDLINE=21218513; PubMed=11319082;  
DOI=10.1128/AEM.67.5.2062-2069.2001;  
RA Morgan J.A.W.; Sergeant M., Ellis D., Ousley M., Jarrett P.;  
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus  
RT PMF1296."  
RL Appl. Environ. Microbiol. 67:2062-2069 (2001).  
DR EMBL; AJ308438; CAC38404.1; -; Genomic DNA.  
SQ SEQUENCE 2538 AA; 283995 MW; 3A7A98CF54FD756E CRC64;

Query Match 95.0%; Score 12396; DB 2; Length 2538;  
Best Local Similarity 95.1%; Pred. No. 0;  
Matches 2424; Conservative 31; Mismatches 71; Indels 24; Gaps 3;  
QY 1 MYSTAVILNKISPTRDQTMTLADLYLSFSELKIPDDQLSMGEARHLYHETIEQKNN 60  
DB 1 MYSTAVILNKISPTRDQTMTLADLYLSFSELKIPDDQLSMGEARHLYHETIEQKNN 60  
QY 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYDMFGARSSSVFKPGSVASMFSPAGYLT 120  
DB 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYDMFGARSSSVFKPGSVASMFSPAGYLT 120  
QY 121 ELYREAKDLHFSSAYHLDNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180  
DB 121 ELYREAKDLHFSSAYHLDNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180  
QY 181 SDALMESLSTYRAIDTPPHQYETIROQVIMTHDSTLSALSNPEVMQAGASLAILA 240  
DB 181 SDALMESLSTYRAIDTPPHQYETIROQVIMTHDSTLSALSNPEVMQAGASLAILA 240  
QY 241 NISPELVNLTETETKNADALFAQNSENITPENFASQSWIAKYKYLGLSEVQKYLQWL 300  
DB 241 NISPELVNLTETETKNADALFAQNSENITPENFASQSWIAKYKYLGLSEVQKYLQWL 300  
QY 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKITRVKTDYDDYDKNINYFDLMYEGNQQFFIR 360  
DB 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKITRVKTDYDDYDKNINYFDLMYEGNQQFFIC 360  
QY 361 ANFKVSRFGATLRKAGPSGIVCSLGPLTANTNFKSNLYLSNDSYKNGKVKLYAVRY 420



RESULT 2	
ID	QBKSP0_KENNE
ID	QBKSP0_XENNE PRELIMINARY; PRT; 2524 AA.
AC	QBKSP0;
DT	01-OCT-2002 (T-EMBLrel. 22, Created)
DT	01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT	01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE	Insecticidal toxin A2.
OS	Xenorhabdus nematophilus.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Xenorhabdus.
OX	NCBI_TaxID=628;
RN	[1]
RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=BP;
RA	Cui L., Qiu L., Pang Y., Yan X., Pang Y.;
RL	Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR	EMBL: AY282763; AM7729.2; ; Genomic DNA.
SQ	SEQUENCE 2524 AA; 282790 MW; B6B6281A9B0B3530 CRC64;
 Query Match 92.8%; Score 12098; DB 2; Length 2524; Best Local Similarity 93.7%; Pred. No. 0; Matches 2379; Conservative 48; Mismatches 95; Indels 18; Gaps 5;	
QY	1 MYSTAVLLNKISPTRDGTMTLADLOYLFSFSELRKIFDDQLSWGGRHLHYHETIQKKNN 60
DB	1 MYSTAVLLNKISPTRDGTMTLADLOYLFSFSELRKIFDDQLSWGGRHLHYHETIQKKNN 60
QY	61 RLLEARIFFTRANPQLSGAIRLGIERDSVSRSDYDEMFGARSSSFVKPGSVASMFSPAGYLT 120
DB	61 RLLEARIFFTRANPQLSGAIRLGIERDSVSRSDYDEMFGARSSSFVKPGSVASMFSPAGYLT 120
QY	121 ELYREAKDLHFSSAYHLNRPPDLADLTLSQSNMDTEISTLSNELLEHHITKKTGGD 180
DB	121 ELYREAKDLHFSSAYHLNRPPDLADLTLSQSNMDTEISTLSNELLEHHITKKTGGD 180
QY	181 SDALMESLSYRQAIDTPHPQYETIRQIVIMTHDSTLSALSARNPEVMGQAGSALLAILA 240
DB	181 SDALMESLSYRQAIDTPHPQYETIRQIVIMTHDSTLSALSARNPEVMGQAGSALLAILA 240
QY	241 NISPELYNILTEITEKNADALPAQNFSENITPENFASQSQTIAKYGLSEVQKYLGM 300
DB	241 NISPELYNILTEITEKNADALPAQNFSENITPENFASQSQTIAKYGLSEVQKYLGM 300
QY	301 ONGYSYSTAYVDNIISTGLVNNESSKLEYKITRVKTDYDDYDKNINYPDLMEGNQPFIR 360
DB	301 ONGYSYSTAYVDNIISTGLVNNESSKLEYKITRVKTDYDDYDKNINYPDLMEGNQPFIR 360
QY	361 ANFKVSREFGATLRKNAGPSGVIGSLGPLIANTWPKNSYLNISDSSEYKNGVKIYARY 420
DB	361 ANFKVSREFGATLRKNAGPSGVIGSLGPLIANTWPKNSYLNISDSSEYKNGVKIYARY 420
QY	421 TSSTSATNOGGGIPTFESYPLTFIPALKNAIKLARLCITGLSPNELQTIIVRSNDAQIIND 480
DB	421 TSSTSATNOGGGIPTFESYPLTFIPALKNAIKLARLCITGLSPNELQTIIVRSNDAQIIND 480
QY	481 SVLTKVFPYTLFYSHRYVALSFDQAQVLNGSVINOYADDSDSVSHFNRLFWNTPLKGKIPEAD 540
DB	481 SVLTKVFPYTLFYSHRYVALSFDQAQVLNGSVINOYADDSDSVSHFNRLFWNTPLKGKIPEAD 540
QY	541 GNTVTSIDPDEEQSTFARSALMRGLGVNSGELYQLGKLACGLAQDAQNTITLSVFVISSLYRL 600
DB	541 GNTVTSIDPDEEQSTFARSALMRGLGVNSGELYQLGKLACGLAQDAQNTITLSVFVISSLYRL 600
QY	601 TLLARVHQVTNVELCMYGLSPNGKTTASLSGBELPRLVILWLYQVOWLTAEAITTEAI 660
DB	601 TLLARVHQVTNVELCMYGLSPNGKTTASLSGBELPRLVILWLYQVOWLTAEAITTEAI 660







RA	Waterfield N.R., Daborn P.J., ffrench-Constant R.H.;				
RT	"Genomic islands in Photobacterium";				
RL	Trends Microbiol. 10:541-545(2002).				
DR	EMBL; AF188483; AAF05542.1; -; Genomic DNA.				
DR	EMBL; AF346500; AAL18486.1; -; Genomic DNA.				
SQ	SEQUENCE 2516 AA; 282953 NW; 8250A0650B614B99 CRC64;				
	Query Match 38.7%; Score 5043; DB 2; Length 2516;				
	Best Local Similarity 42.6%; Pred. No. 3.7e-259;				
	Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;				
Qy	22 LADQVLSFSLRIRKIFDDQLSGWGBARHLYTHETIQKNNRLLLEARIETFRANPOLSGAIRL 81	Db	885	QQLNVAPOQGVSAVLGLDYIQSMKET----	PTYAQWENAAAGVLTAGLNSQQANTLHAFIDE 940
Db	21 LTDLSHSFNEFRQVSEHLWSETHDYHDAQQAQKDNLYEARILKRAPOLQONAVHL 80	Qy	941	RLSGLCNWFLANIOPEGVSLHSDDDLYSYFLDNOYSSAIAKTRTLEAETAGIOLYNRA 1000	
Qy	82 GIERDSVSR-SYDMFGARSSSVKPGSVASMPFAGVLTLYREAKDLHPSSAYHLDN 140	Db	941	SRSAAALSTYIRQVAKAAAIAKGRDDLYQYLLDQNSAAIKTRTLEAETAGIOLYNRA 1000	
Db	81 AILAPNAELIGYNNQFSGRASQYVAPGTVSMFSPAAYLTLYREARNLHSDSVYLD 140	Qy	1001	LNRIEPNARADVSTROQFTDW-TVNNRYSTGGVSRVLYVYPENYIDPTQRTGQTRMDEL 1059	
Qy	141 RRPDLADLTLSQNMDEITSLTSLNELLLEHI-TRKTGSDSALMESLTYRQAITPY 199	Db	1001	LENVEENANGSVLSRQFFIDWDKYNKYSTWAGVSQVLYVYPENYIDPTMRTGQTRMDEL 1060	
Db	141 RRPDLKSMALSQNMDELSTLSUSNELLLESIKTESKLENYTKVMEMLSTFRPSGATPY 200	Qy	1060	LENISQKLSRDVEDAFKTYLFRFETVADLVKVSAYHDNNSVNTGLTWFGQTRMDEL 1119	
Qy	200 HQPYETTRQVMTHTDSTLSALSRPEVMOQAEASILLAILANISPELYNILLTEITEKNA 259	Db	1061	LQSVSQSLNADTVEDAFMSYLSFEQVAMLVISAYHDNINNDQGLTYTIGLSETDAGE 1120	
Db	201 HDAYENREVITQDPGLEQLNASPAIAGLMHQASLLIGINASISPELFNILLTEITEGNA 260	Qy	1120	YYRWNVDIRMQAGELAAWAKWKIDTAVNPVKDAIRPVIPERLHLIIVWEKEEVAK- 1178	
Qy	260 DALFAQNFSENITPENFASQSWIAKYKYLELSEVQKYLGMQN-GYSDSTSAVYDNI 318	Db	1121	YYRVSVDHSKPNQKFAANAWSEMHLKIDCPINPYKSTIRPVYKSRLLYLLWLEQETIKQ 1180	
Db	261 EELYKKNFG-NIEPASLAMPKYLKRYNLSDEELSQFGRKASNPQGBYSN---NQLITP 316	Qy	1179	--NCTD--PVETDYDRFTKLAFRLHSDGSWAPSVDITTOVEAVTDDKPPDTERLALAASG 1234	
Qy	319 LVVNNESKLEAYKTR-VKTDYDKNINYPDMYEGNQQPFIANPKVSRFEGAT----- 372	Db	1181	TGSKDGYQYETDYRYELKLAHRYDGTWNTPTTFDYNKKISEL--KLEKNRAPGLYCAG 1238	
Db	317 VNNSDGTGVKYRITREYTNAYQMDVELFP--FGGEN---YRUDYFKPNFYNASYLSIK 371	Qy	1235	FOGEDTLVLYFKYKGSYDFGSGSNKNAVGMTIYDGSFKKMENTALSRYSQLKNFTDII 1294	
Qy	373 -----LRKNAGPSGIVGSLGSLIANTNPKSNVLSNISSEYKNGVKIYAYRYTGSTS 425	Db	1239	YQGEDTLVLYFYNQDITLDSY--KNASMOGLYIFADMASKDMTPEQSNVYRD--NSYQQF 1294	
Db	372 LNDKRELVRTEGAQ-----VNIEYSANITLNTAD--LSQPFEGLEFVLPSSG 418	Qy	1295	HTQGNDLVRKASVRFPAQDFEVPASLANWS--AIGDSDLTVMENGINIQTISKYSSDNLAI 1352	
Qy	426 ATNQGGIFTFESYPLTFALKAKAIRLCITSGLSNPELQITVRSNDAQIINDSVLT 485	Db	1295	DTNN---VRVNNRYAEDYIPSSVSRKDYGMWDYLSVMYNGDITPINYKAASSDLKI 1351	
Db	419 WA-YAAAKFTVEEYNOYSLFLKAKAIRLSRATESLPTILEGIVRSVNLQDINTDVLGK 477	Qy	1353	-----TLHNAFTVRYDGSNVIRNKQISAMKLTGVGDKSQVGNAPFIANTVKHYGY 1405	
Qy	486 VFYTLFYSRYALSFDDAOLVNGSVINQYADDDSVSHENRLEFNTPLKGLFEADGTVS 545	Db	1352	YISPKLRIIING-----YEQG-----KRNQCNLMNKG-----KLGDKFIV-----Y 1388	
Db	478 VFLTKYMQRYAHAEATILCNAPIQRSYDNPQSFQDRLENTPLNGQYFSTGDEID 537	Qy	1406	SDLG-GBITYVYNKTKNY-IASVQGLMADYTRRLILTPVENNYIARLFEP-PPSPNTIL 1462	
Qy	546 IDPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLAQNTITLSVFVLSLYRLTLAR 605	Db	1389	TSLGVNPNNSGNKLMFPVYQYSGNTSGLNQR--LLFHRDITTPSKVEAMIPGAKRSLT 1446	
Db	538 LN-SGSTGDWRKTLKRAFNIDDVSLFRLKIITDHDNKGKIKNNKLSNLYIGKLAD 596	Qy	1463	NTVFTVGS-----NKTSDFKCSYAVDGNNSOGFOIFSSYOSSGWLDDITGNNNDIK 1515	
Qy	606 VHQTLVNELCMLYGLSPNGKTTAS-LSSGELPLRVLVQVTOWLTEAEITTBAILWC 664	Db	1447	NQNAAGDDYATDNLNPKDQLKQYIFMTD---SKG-----TATDVSGVPEINTALSPAKVQ 1499	
Db	597 IHQTLIDELDLL-IAGEGKTNLSAISDKQLATLRKANTITSLWLTQKMSVPLFMT 655	Qy	1516	ITVMAGSKTHTFTASDHASLPANSFDMPTFKPLEIDASSLAFTNNAIPLDIVFTKA 1575	
Qy	665 TPESGMSIPESINLNNRPSI---SEDMAQSHNRELQAEILAPFAATLHLASPDWAR 721	Db	1500	IIVKAGKEQTFADKDVSTQSPSFDENYQFNALIDGSLFINNSASIDVTFATA 1559	
Db	656 STSYNKTLTPEIKNLLTVYHGLQFGDKAD-----LHVMAPYIAATLQLSSENVAH 709	Qy	1576	KDGRVLG-----KIKQTLSSKRVNVPEDIIFLRTHSGAQVQMLGVTRIRINTLLASQLV 1631	
Qy	722 YILLNTNLRPGGLDIAGFMT-----LVLESINANETQ-----LVQFCHVMAQLSLS 770	Db	1560	EDGRKLYGSFSPVTLKV-----STDNALTLAHNENGAQYMQWQSYRTRTLTLFARQLV 1614	
Db	710 SVLLWADKLQPD-----GAMTAEFWDWLNKTYTPGSSEAVEQEHVVOYQALQLEMV 765	Qy	1632	SRANTGIDTILTWETQRLPEPPLGEGFFANFVLPKYPAEHGDERRWFKIHIGNVGGTGR 1691	
Qy	771 VQTLRLSEALSLVIVISGFAVLGAKNPAGQGHNTDTLPSLRFHQWINGLNGPSGDTLDM 830	Db	1615	ARATGTGIDTILSMETQIQBPQKGFYATFVIPPYNLSTHGDERRWFKLYIKHVVDNNS- 1673	
Db	766 YHSTGINENAPRLFVTKP-EMFGAATGAAPAHDAISLIMLTRFADWVNALGEKASSVLA 824	Qy	1692	QPYVSGMLSDTSETSMTLF-----VPYAEGYMHGVRGLGVQYKITYDNTWESAFFYD 1746	
Qy	831 LRQOTLTADRLASVMGLDISWTOAMVSA-----GVNQLQCODINTVLOWIDVA 880	Db	1674	HIYSGQLTDTN-INIITLFIPLDDVPLNQDYH-----AKVYMTFKKSPSDGTWGWGPHFVRD 1728	
Db	825 FEANSLTAEQADAMNDANLLQASIQAOQHQLHPVPTPENAFSCWTSINTILLQWNV 884	Qy	1747	ETKQOQFVLINDADHSDGMOQGIKVKIKKYGFLNVSITATGYSAPMDFNSASALYIYELF 1806	
Qy	881 SALHTMPSVIRTLVNIKRYVTALNKAESNLPDSEWQTLAENNEAGLSLTOQAQTLADYAE 940	Db	1729	D--KGIVTIN-----PKSILTHPESVNLNNIS-----SEPMDFSGANSYFWELF 1772	

```
QY 1984 NPALTDYQWTLRLRLNLRHNLSDGQPLSLAIYAETDPKALITSMVQASGGSAVLPG 2043
D 1953 NEVMWNYQWTLAQRYVYNLRHNLSDGQPLSLAIYAETDPKALITSMVQASGGSAVLPG 2012
QY 2044 TSLYRFPVWLERTNLVAQLTQFTSLSSWAHDDADELTLLLOQWMLATQSIROQ 2103
D 2013 FMSLWRFPFPHLENARGMSVQLTQFTSLSSWAHDDADELTLLLOQWMLATQSIROQ 2072
QY 2104 RTVDEVDADIIVLAERSSAONRLKRYOQLYDDEINHGEORAMSLDDAAAGSLAGQVLS 2163
D 2073 KTIIELDKTVLEKSKAGASRPDSYKGLVDENINAGENQANTLRASAGLITAVQASR 2132
QY 2164 IAEGVADLVPNVFLGACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRRRRQEW 2223
D 2133 LAGAAADLVPNIFGPAGGSRWGAALRASAVMSLSATASQVSADKISRSEAYRRRRQEW 2192
QY 2224 EIORDNAGEVQOMDAQLESKIRREAAQOMVEQETOQAOQLELLOQWMLATQSIROQ 2283
D 2193 EIQRNABAEIKQIDAQLKSLAVREAAVLQKTSKTOOQTSQLEAFLOKFSNQALYN 2252
QY 2284 WMRGKLSIYYQFDFLTQSCFLMAQEARLRELTDNGVTFIRGGAWNGTTAGIMAGETLL 2343
D 2253 WLRGLAIIYQFDFLTQSCFLMAQEARLRELTDNGVTFIRGGAWNGTTAGIMAGETLL 2312
QY 2344 NLAEMEKVWLRDERALEVTRTVSLAOFYQALSSDN--FNLTEKLTQFLREGKNGVAGS 2401
D 2313 SLAQMEDAHLKRDKEALREVERTVSLAOFYQALSSDN--FNLTEKLTQFLREGKNGVAGS 2372
QY 2402 NELKL-----SNRQIEASVRLSDKIFSDYPESIGNTRQLQKOVSTLPAVGPVEDIRAV 2456
D 2373 NTLAFAGACTDKTSLQASVSFADLKIRIDYFASLGKIRRIKQISVTLPALGYPQDVQAI 2432
QY 2457 LNYGSSIVMPGCSAIALSHGVNDGQFMDPNDISRYLPPFGISVNDGSLTLFPDPAT- 2515
D 2433 LSYGDKAGLANGCEALASHGVNDGQFMDPNDISRYLPPFGISVNDGSLTLFPDPAT- 2515
QY 2516 ---DRQKALLESLSDIILHRYTIR 2537
D 2492 PEKGQATMLKTLNDIILHRYTIK 2516

RESULT 5
ID O85160 PHOLU PRELIMINARY; PRT; 2504 AA.
AC O85160
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-FEB-2005 (TRENBLrel. 29, Last annotation update)
DE Insecticidal toxin complex protein TcBa (Toxin complex protein).
GN Name=tcbA;
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_TaxID=29488;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photobacterium luminescens."
RL Science 280:2129-2132(1998).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
```

```
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047457; AAC38627.1; -; Genomic DNA.
DR EMBL; AF346438; AAL18460.1; -; Genomic DNA.
DR InterPro; IPR002345; Lipocalin.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
SQ SEQUENCE 2504 AA; 829581693D1047CE CRC64;

Query Match 38.2%; Score 4986.5; DB 2; Length 2504;
Best Local Similarity 42.3%; Pred. No. 3.7e-256;
Matches 1110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;

QY 9 NKISPTRDG-----QTTWTLADLQYLSFSELRKIFDDQLSGEARRHLYHETIQKNNRL 63
D 3 NSLSTIDTICQKLQLTCPABIALYPPDTPREKTRGMVWMEAKRIYEAQEQDNRLLH 62
QY 64 EARIPTANPOLSGAIRLGIERD--SVRSYDEMFGARSSVFKPGSVASMFSPAGLYTEL 122
D 63 EKRIFAVANPLLNKAVRLGTRQMLGFTQGYSDLFGRADNAAAFSGVASMFSPAAYLT 122
QY 123 YREAKDLHFSSAYHLNRRPDLADLTLSQSNMDEITSLTSLNELLEHITRTKGTGSD 182
D 123 YREAKDLHFSSAYHLNRRPDLADLTLSQSNMDEITSLTSLNELLEHITRTKGTGSD 182
QY 183 ALMESLSTYRQAITPYHQPYETIRQVIMTHDSTLSALSRNPVGMQAGASLLAILANI 242
D 183 EVMDMLSTYLSGETPYHAYVETVREIVHERDPGRHLSQAPIVAALDPTVLIGISSHI 242
QY 243 SPELYNLTBEITKNN---ADALFAQNFSENIPTENFASQSWIAKYIYGLSEVQYKLM 299
D 243 SPELYNLTBEITKNN---ADALFAQNFSENIPTENFASQSWIAKYIYGLSEVQYKLM 299
QY 300 LQN--GYSDTSAYVDNISTGLVWNESEKLEAYKTRVKTDDYDKNINFDLMYEGNNOFF 358
D 302 LSHVGYIS-----SDILVPLVDGVGKWEVVRTRTPSDNITQTNIELYPPQGGDYL 354
QY 359 IRANFKVSRFEGAT---LRKNAGPSGIVGSLSGP---LIANTFNKSNYLSNISDSYKNG 412
D 355 IKYN--LSNSFLDGFYLYQYKGSADWTEIAHNPVPMVINOQYEQATIKRSDSD--NI 410
QY 413 VKIYAYRTSSTATNOGGGIFTPESTPLTTFALKLNKAIKCLTSLGSLSPNELQTVRSD 472
D 411 LSLGLQRWHS--GYSNFAAANKIDQSPKAFLLKMKAKAIRLLKATGSLFATLERIVDSV 468
QY 473 NAOGIINDSVLTQVYFTLFTYSHYALSPDDAQVINGSVINOYADDSDVSHENRNPNTPL 532
D 469 NSTKSIITVEVLNKRVRKFIYDRIGISETHAAILANINISQAVGNQSLSQEQLFNHPL 528
QY 533 KGKIFE-ADGNT-----VSIDPDEQSTPARGALMRGLGVNSGELYQL-----GKLAG 579
D 529 NGIRVEISEDNSKHLNPNLKPDSCTDDQKAVLKRAFOVNASELYQMLLITDRKEDG 588
QY 580 VL--DAONTITLTVFVSSLYRLTLARVHOLTVELCMYGLSPFNKGTATSSSGSLP 637
D 589 VIKNNLEN-----LSDLVLSLAQIHLNLTAEHLNLTLCVGYGDTNITQITDDNLA 640
QY 638 RLVIWLVQVQWLTAEAITTEAIVLCTPESFGNISPEISNL-----LNNLRPSISED 690
D 641 KIVTLTWITQWLTQWTVTDLMTATVSTLTTPESINLTATLSLTGKESLIGED 700
QY 691 MAQSHNRELOAEIILAFPIATLHLASPDARYIILWTDNLRPGGLDIAGFMVLVKESLN 750
D 701 LKRA-----MAPCFTSALHLSQEVAVDILLWIDQIQPAQITVDGFW-----EVQ 746
QY 751 ANETT-QLVQFCHMAQLSLSVQTLRLSEALSVLVISGFAVLGAKNQAPQACHNIDTIFS 809
D 747 TTPTSKLVITPAQVLAQLSLIYRRIGLSELSLIVTQSSILLVAGKS--ILDHGLLTMA 804
QY 810 LYREHONGINGNPGSDTLDMLRQTLTADRLASVMGLDIDSMVTOA--MVSAGVNOLOCW 867
D 805 LEGPHTWVNGLGHQASLILAAKQDGLTVDVAQMMKESLLOQAAANQVKEKDTIKLTSW 864
```







Db 892 DWDKYNKYSTWAGVSQLVYYPENYIDPTMRIGOTKMMDTLLQSVSQSLNADTVEDAPK 951  
Qy 1079 TYLTFETVADLKVVSAVHDNNSNTGLTWFVQOTRENLPEYRNVDISMAQAGELAN 1138  
Db 952 SYLTSFEQVANLEVISAVHDNNDNDQGLTYFGNSKTEVQYVRSVDHSEKDFKPAAN 1011  
Qy 1139 AKWETKIDTAVNPYKDAIRPVI FRERLHLIWEKEBAKNGTDPVETD - RFTLKAPL 1197  
Db 1012 AWSEMHKIDCAINPYQSTIRPVIYKSLYLIWLEQKETAKQEDNKVTTDYHYELKLAHI 1071  
Qy 1198 RHDSWSAPWSYDITTOVEAVTDKPDTERLALASGFGEDTLLAVFYVTKGKSYDFGG 1257  
Db 1072 RYDGTWNPITFDVDEKILALELTK - SOAPGLYCAGYQGEDTLLIMFYRKKEKLDY - 1127  
Qy 1258 SNKNVAGMTIYDGGFKKMENTALSRYSQKNTFDIHTQGNLDRKASVYFAQDFEYPA 1317  
Db 1128 KTAPOGFVIFSDMGSKDWTNEQNSYRDNQVYTHFTDTSNDSVIR - INNRVAEDYEIPS 1186  
Qy 1318 SLNMGSA - IGDSLTVMENGINPOITSKYSDNLAITLHNAAFVTVYDGGNVRNKQI 1375  
Db 1187 LINSNSHDWGEYNLSQVYGGNI - VINYKVTSDNLKIYI - 1224  
Qy 1376 SAMKLTGV - DGKSQGNAPFIANTVKKHYG - YSDLG - GPTIVY 1415  
Db 1225 -SPKLRIIHDGKE - GRERIQSLNLIKYGKLGDKFIIYTSGLINPNNSNRPMFYPVYQ 1281  
Qy 1416 NKTNYIASVOCHLMNADYTRRLILTPVENNYARLFBF - PPSNPNTILNTVTVGSG - 1470  
Db 1282 NGNTSGLA - QGRLL - FHR - DTSYSSKVAAWIFCAGSLINENANIGDDCAE 1329  
Qy 1471 ---NKTSPFKCSYAVDGNNSQGFQIFSSYSSGWMLDIDTGINNTDIKITVWAGSKTHTF 1527  
Db 1330 DSVNKPDDLKQVIYMTD - SKG - TATDVSGPVDINTAISSEKQVITIKAG - KEYSL 1381  
Qy 1528 TASDHASLSPANSFAMPYTFKPLEIDASSLAFTNNIAPLDIVFETKAKDGRVLG - KIKQ 1586  
Db 1382 TANKDVSQVSPFSFEEMCYQFNALBIDGSLNFTNSASIDVTFALADDGRKLGYEIPN 1441  
Qy 1587 TLSVKRVNVPEDILFLRETHSGAQMOLGVVRIKRLNTLLASQLSVRANTGIDTILWET 1646  
Db 1442 IPVIOKV - KYDNALTLPHDENGAYQMWGAYRIRLNTLFRQLVERANTGIDTILNET 1499  
Qy 1647 QRLPEPPLEGFFANFVLPKYDPAABHDERMPKHIHGVNGNTGQPYSGMLSTSETS 1706  
Db 1500 QNIQEPWGMIGAYIELIDKYNPDTHGTGKFKIYIGDIFRAGDHPIYQGALSIDITQT 1559  
Qy 1707 MTLFVPYAEQYMH - GYRLGVGYQKIYDNTWESAPFYFETKQOQVFLINDAHDGWT 1765  
Db 1560 VKLFLPRVDNAYGNKNLYVVAAYQKV - ETNFIRFVKEDNNKPAT 1603  
Qy 1766 QQGIKNTKKYKGFNLVSIATGYSAPMDFNASALYYWELFYTPMCMCFQRLLOEKQFDE 1825  
Db 1604 FDTYKN - GTPGLASARVITQVSEPMDFSGNSLYFWELFYTPMVAQRLHQNDFE 1662  
Qy 1826 ATQWNYNYPAGYVINGEIPWICNCRPLETTSWNNANPLDAIDPDVAQNDPMHYKIA 1885  
Db 1663 ANRWLKYVWSPSGYVIRGQIKNYHNVRPLENTSWNSDPLDSVDPDAVAGHPMHYKVA 1722  
Qy 1886 TFMRLLDOLILRGDMAYRELTALNEAKMYVTRLELLGDEPEDYGSQWAAPSLSGA 1945  
Db 1723 TFMRLTLLMARGDHAYRQLERDTLINEAKMWYMQALHLLGNKPYLPVLSVWMDPRLDNA 1782  
Qy 1946 SQTQAAVQOQDLMGRGVSKN - LRTANSVLGLFLPEYNPALYDYQWTLRLRLNLRH 2003  
Db 1783 ATTQKHAHAYITSRQGTQTPALLIRSANITLDFLPQINDVMSYNWKLRLNLRH 1842  
Qy 2004 NLSIDQPLSLAIYAEPDTPKALLTSMVQAQSGGSVILPGLTSLYRFPVPMLETRNLVAQ 2063  
Db 1843 NLSIDQPLHLPIYATPADPALLSAAVATSQGGKLPESFISLWRFPHMLNARSMTQ 1902  
Qy 2064 LTQFTSLLSWAHDADDELITLLLOQGMELATQIRIQORTVDRVDADIIVLAERSRA 2123  
Db 1903 LIQFQSTLQNIETIERQDAESLALLQNAKELITLTSIQDKTIEIDAETVLEKSKAGA 1962

Qy 2124 QNRLEKYOQLYDEEDINHGEQBAMSLDDAAAGQSLAGQV - LSIABGVADLVNPFGLACG 2181  
Db 1963 KSRFDNYSKLYDEEDVWAGERQALDM - RIASQITSGLKGLHMAAAALEMVPNIYFVAVG 2020  
Qy 2182 GSRWGAALRASAVMSLSATASQYSADKISSEAYRRROEWEIQORDNADGEVKQMDAOL 2241  
Db 2021 GTRYGAIANATAIOGGIAAEGLLTEAEKVYSEIWRRRRQBEWEIORNNAEAMKQIDAO 2080  
Qy 2242 ESLKIRRAAQQVQYEQTOQAHTQAOALELQKFTNKALYSWMRGKLSAIYQYFDFLTQ 2301  
Db 2081 KSLTVRREAAVLQKTGLTKTQEQOQAFLQRFESNOALYNWLRGLAAIYFQYDVLV 2140  
Qy 2302 SFCMAQALRELTNDNGVTFIRGANGTTFAGLMAGETLLILNLAEMKVMLEDERALE 2361  
Db 2141 ARCLMAEQAYRWETNDSARFIKPGAQGTQVAGLAGETLMLNLAQMDAHLKQBRAL 2200  
Qy 2362 VTRTVSLAOFYQALSSDNFNLTQFLRSGK - GNVGASGNELKLSNRQTEASVRLSD 2420  
Db 2201 VERTVSLAQQVYSGEKSFAUKKIALLOQDKETSAQDNGNQLKJTNNTUSATUTLQ 2260  
Qy 2421 KIFSDYPE - SLGNTRQLQKQSVTLPALVPYVEDIRAVLNTYGGSI - MPRGCSAIALSHG 2477  
Db 2261 KLKDDYPEMQLKTRIKQISVSLPALLGYQVQAVLSYGGDATGLAKGKALAVSHG 2320  
Qy 2478 VDSGQFMLDPNDSRYLPFEGISVNDSSLTSLSPDPDATROKALLESILHIRYTI 2536  
Db 2321 LNDNGQFOLDPNDGKFLPFEGIDINDKGTFTLSPNAASKQKNIILQMLTDIILHIRYTI 2379

## RESULT 8

Q7N7Y6 PHOLL  
ID Q7N7Y6 PHOLL PRELIMINARY; PRT; 2378 AA.  
AC Q7N7Y6;  
DT 01-MAR-2004 (TreeRel. 26, Created)  
DT 01-MAR-2004 (TreeRel. 26, Last sequence update)  
DT 01-MAR-2004 (TreeRel. 26, Last annotation update)  
DE Insecticidal toxin complex protein TcdA4.  
GN Name=tcdA4; OrderedLocusNames=plu0965;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
EX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bocs S., Bourdeau-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derosé R., Derzelle S., Freyssinet G., Gaudriault S.,  
RA Médigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunst F.,  
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
RT luminescens";  
RL Nat. Biotechnol. 21:1307-1313 (2003).  
DR EMBL; BX571862; CAE13260.1; -; Genomic\_DNA.  
DR Photoflist; plu0965; -.  
KW Complete proteome.  
SQ SEQUENCE 2378 AA; 271176 MW; 09834DDDBF1D4C6B CRC64;

Query Match 36.5%; Score 4763.5; DB 2; Length 2378;  
Best Local Similarity 40.5%; Pred. NO. 2.7e-244;  
Matches 1047; Conservative 447; Mismatches 795; Indels 293; Gaps 47;

Qy 22 LADLYQSFSELRKTFDDQLSGEARHLYHETIEOKNNRLLERIFTRANPQLSGARL 81  
Db 21 LTDICHVSFNRQVSDHLSWSENRNRYRDAQQKQENQLYEASILKRANPQLQNAHL 80  
Qy 82 GIERDSVS - RSYDEMFGARSSSFVKPGSVASMFSPAGYLTETLYREAKDLHFSSAYHLDN 140  
Db 81 GITLPHAELOQYNNEFGGASQYVAPGSSVMSFPAAYLTETLYREARNLHASDSVYHLDQ 140  
Qy 141 RPLADLTLSQSNMDSITSLTSLNELLLEHI - TRTKTGGSDA - LMESLSTYRAIDTP 198



Db	141	RRPDLQSMTLNQNMDELSTLSLSNEILLEGKANKQPNLNDTKVMAMLSLTPRPGSTIP	200
Qy	199	YHQPETYIRQVIMTHDSTLSALSNNPVMQOAGASILLATIANISPELYNMLTEETEK	258
Db	201	YHDAYENVRKAIQLODPOLQFOQSPAVAGLMHQASLLGINNSISPELNFILTEETAN	260
Qy	259	ADALFAQNPENITPENFASQSWIAKYGL---ELSE-VOKYLMQLONGYSDSTSAYVDN	314
Db	261	AEDVYKKNFGD-IDPTWLANPEYLKSYNLSDELSQFIRKY-----	301
Qy	315	ISTGLVNNESKLEAYKITRVKTDYDKNINYFDLMTYEGNQPFFIRANFKVSRFGATLR	374
Db	302	-----ENNE-----	305
Qy	375	KNAGPSIGVCSLGSPLIANTNPKSNYLSNISDSEYKNGKVIYAYRYTSSATSATNOGGIP	434
Db	306	-----LNTQ-----	309
Qy	435	TFESYPLTIPALKINKAIRLCLTSGLSPELQTVIRSDNAQGIINDSVKTVFTLPSYH	494
Db	310	-----KIHLKINKILLRSVNLFPVKDELIIPEQS-----INPSVLGKIFLVKYMQ	358
Qy	495	RYALSPDDAQVLNGSVINQYADDDSHFNRLFNTPELKGKIFPADGNTVSIIDPDE-EQS	553
Db	359	KYNIDTETALILCNNSISQSYRNQPSQFRLFNTPELNGOCFVTEDTNIDLNLNSTDTD	418
Qy	554	TPARSALMRGLGVNSGBELYQGLKGLAVLDAQNTITLSVFVSIYRLTILARVHOLTVNE	613
Db	419	NWHTKVLKRAFNVDDIISLYLLHIAHNEKDGKIANNIKNLSNLYMTKJLADTQHLTIDE	478
Qy	614	LCMLYGLSPNGKTTASLSSGELPRLVIWVQVOTWLTEABITTEATWLCTPEFSGNIS	673
Db	479	LYLLMTIGEDKRDLYIIDDKLEKLNRLDLSNWLRTQKWSYQVLFMTTNYDKTLT	538
Qy	674	PEISNLLNLRPSISEDMAQSHNRLEQAEILAPPIAATHLASDPMARYILLMTDNLRLPG	733
Db	539	PEIQNLLNTVYNGLON--FDENKKOLLTISPIYAATLQPSENVVAHSILLWADKIKPS	595
Qy	734	GLDIA-----GFWTLVLKESLNANETTQLVQFCHVMAQLSUSVOTLRLSEALS	782
Db	596	EKKITTEKFDWLQNRG-----ILELSKPPEMQEQIIQYCHCLAQTLMIYSESGINENAFR	651
Qy	783	VLVISGPAVLGAKNOP--AGQHNIDTLFSLYRPHOMINGLGNPGSDTLOMLRQOULTAD	839
Db	652	LFIEQPTIFIGISDEPKATPAHNAPLLIILTRFANWNTLGEKASPLITAFENKILTAE	710
Qy	840	RLASVMGLDISMTVQAMVSA-----GVNQLQCWDQINTVLQIDIVASALHTMPSVIRT	892
Db	711	QLANAMNLDAILLQASIQAGNYQKVTKENTFSNWSQSIDIIQLQTNVANDFNISPOGISS	770
Qy	893	LVNIRYVTLNKAESNIPSWDEWQTLAENWEAGLSTQOATLADYTAERLSSVLGNWFLA	952
Db	771	LIALDYIQPSOKT----PTYSQWENAAVLTAGLDTQOmntTHAFLEDSRSTALSNIYIS	826
Qy	953	NIQPEGVLSHRDDLVSFYELDNOVSSAIKTRTLAEAIAGIOLYINRALNRIBENRADV	1012
Db	827	KIANRAAIIKSRDDLYQLLIDNQVSAIKTRITRAEAIASIQLVNRALENIETHAISDV	886
Qy	1013	STROFFTDW-TVNNRYSTWGSVRLVYYPENYIDPTORIQTORMDELLENISQSLSRD	1071
Db	887	ITROFFIDWKYNKRYSTWAGVSQLAYYPENYIDPTWRIQOTKMMDTLQSVQSQSLNAD	946
Qy	1072	TVEDAFKTYLTRPFTVADLKVVSAYHDVNSNTGLTWFGVQGTRENLEPYTWRNVDIRMQ	1131
Db	947	TVEDAFKSYLTSPFOQANLEVISAYHDVNVNNDQGLTYFIEGHSGTEINQYTWRSYDHSKFN	1006
Qy	1132	AGELAAWAKWETKIDTAVNPYKDAIRPVLFRERHLIIWKEBKVAKNGTD-PVETVDRF	1190
Db	1007	DGKFAANAWSEWHKIDCAINPYQSTIRPVYKSRLYLIWLEQKIAQDKDNKVITDYYY	1066
Qy	1191	TLKLAFRLRHGDSWSPAPSYDITITQVEAVTQDKPDTERRALAAAGSQFQGEDTLLVFFVYKTKG	1250

1067	Db	ELKLAHRYDGTWNTPIAFDVNSKISALELAK--SOAPGLYCAGYGBETLLVMFVRKKE	1120
1251	Qy	SYSDFGSNKNVAGMTIYDGSFKOMENTALSRYSOLKNTFDIIHTHQGNDLRKASYSREA	1310
1125	Db	KLDDY--KTAPMOGFYFSDMSSEKMTNQCSYRDNGYTHPDTNSDTSVIR-INNRYA	1181
1311	Qy	QDFEVPASLNMGSA--IGDDSLTVMGNGINPQITSKYSSDNLAITLHNAAFVRYDGSNG	1368
1182	Db	EDYEIPSLINHNSHDWGEYNLSQVYGGNI-----VINEVASS	1220
1369	Qy	VIRNKQISAMKLTCVDKQSOYGNAPIANTVHKYGGYDLGGPIVY-NKTKNYIASVOG	1427
1221	Db	DLKIKISPKLRIITH-EGKE--GRERIQCDLMKKYG---KLGDKFIIYITNKNKNPNHKSNE	1274
1428	Qy	HLMNADYTRLLILTPVENNYAALFE-----PFPSNPNTILNTVFTGSKNKTISDF	1476
1275	Db	KLIYVVOYR-----NNDKGCLLFVRSNTTGTIRAWFPTTKEEITAT-----TSSN	1320
1477	Qy	KKCSYAVDGNNSOGFOIFSSY-----QSSGWLDIDPTGINNTDIIKITVMAGSKT	1524
1321	Db	QDC--IIDTINN--IDLKAYFYMEDKEGQIATDISGATDINTGISSEKVOIINAG-KE	1375
1525	Qy	HTFTASDHIASLIPANSPDAMPYTFKPLEDASSLAFTNIAPLDIVFETKADGRVLG-K	1583
1376	Db	YSLTANKDVQSPSPSEEMCYOFNALEYIDGSSLNFNTNNSASIDVTFFIALADGRKLGYE	1435
1584	Qy	IKOTLSVKRVYNPEDILFLRETHSGAQYMQLGAVRIRLNTLILASOLVSRANTGIDTIILT	1643
1436	Db	IFNIPVLOKV--KTDNALTLFHDKNAGQAYQWGAIVRLNTLTFARQLVERATTTGIDTILS	1493
1644	Qy	METQRLPEPLPBGFPANFVLPKYDPAEHGDERWFKIHIIGNVGNTGROPYYSGLMSDRTS	1703
1494	Db	METQNIQEPMMWIGAVYIEIVMDKYNTDIHGTKNSFKIMYGDIFKAGDHPFIYQGTLSDIT	1553
1704	Qy	ETSMTLFVPYABGYMHYH--GVRIGCVGYOKITDYNTWTWESAFFYDQKQFVLINADHDS	1762
1554	Db	QTTVKFLFPRVDNAYGNKNLNYAAYAKA-----ETHFIRPVKEDNNKPATFDTTYNN	1607
1763	Qy	GMTQOGIVKNIKKYKGPLNVSITATGYSAPMDFNSASALYYWELFYFTPMCMQFRLIQEKQ	1822
1608	Db	G-----TFPGLASARVLOTSEPMDFSGANSLYFWELFYFTPMVAQRLLEHQN	1656
1823	Qy	PDEATOWINTVYNPAGYIVNGEIAPIWNCRLPEBETTSWNANPLDAIDPDVAQNDDPMY	1882
1657	Db	FDETNRWLKYVWSPSGYIVRGQIKNYRWNVRPPLENTSWNSDPLDSVDPDAVAQDHPMHY	1716
1883	Qy	KIATFMRLLDQLILRGDMAYRELTDRALHAKWXYRTLELIGDEBEDYGSQOWAAPSLS	1942
1717	Db	KVATFMTETLLMARGDHPAYRQJERDITLNEAKWYMQALHLLGDKPFLPLSSAWNPRLD	1776
1943	Qy	GASQTVQAAAYQODLTWMLGRGGYSKN--LRTANSVLGLFLPEYNPALTDYWOLTRLRLFN	2000
1777	Db	NAODTTIQAISYAITTLRGOTQTPALLLRASANTLDTLFLPQINDVWLSYWNKLEURLYN	1836
2001	Qy	LRHNLSDGQPLSLAIYAETPDKALLITSMVQASOGGSVALPCTLSLYRFPVMLERLNL	2060
1837	Db	LRHNLSDGQPLYLPMYATPADFKALLSAAVATSGGGSLLPOPFMSLWTFPHMLENARSM	1896
2061	Qy	VAQLTQRTGSLLSMAEHDDADELTITLLQCGMELATQSIIRIQRTVDEVDADIIVLAESR	2120
1897	Db	VAQLTQFGSTLQNIIEHQDAEALNALLQNAQKELIITLSIQDKTITEEDAEKTVLEKSK	1956
2121	Qy	RSANRLKCYOQLYDEDINHGEORAMSLIDAAAGQSLAGOV--LSIAEGVADILVPNVFGL	2178
1957	Db	AGAQRFDNTSKLYDEDVNAGEQALDM--RIASQISITGLKGLHMAAAALEWPNYIGF	2014
2179	Qy	ACGGSRWGAALRASASVMSLSASQYSDAKISRSFAYRRRRRQEWESIQRDNADGEVKQMD	2238
2015	Db	AVGGTRFGAANAIAIGGGIAEAGGLLIEAEKVQSEIWRRRRRQEWESIQRNNAEAMKQID	2074
2239	Qy	AQLESKIRREAAQOMQVEYQETQOATQOALELLOKFTFNKALYSNWRGKLSAIYYQFFD	2298
2075	Db	AQOQSLTVREAAVLQKTSUKTQOQOQALQALFQKFNQALYNWLRGLLAIYFQFYD	2134



Matches	1056;	Conservative	411;	Mismatches	810;	Indels	355;	Gaps	55
Qy	1	MYSTAVLLNKISPTR-----DGO	TMTADLOVLSFSELRKIFDDQL	SWGEARHLYHETI	54				
Db	6	MYNIDDILEKVNAPRLSEENDTAVTL	DLFSRSPPEVKKTTGDSL	SLMGCVCVLSQAQ	65				
Qy	55	EOKNNRLLIARIFTRANPOLSGAIR	LGIHTERDSVRSYDEMFCGARS	SSFKVPCGSVASMFS	114				
Db	66	HEOKENRLTESRILARANPLVNAVR	LGRIRQAAGSRSYDDWFGSR	ADRFARPQGSVASMFS	125				
Qy	115	PAGVYELVREAKDLHFSSAYHLNRR	PLADLTLQSQNMDETISTLTLS	NELLLEHIT	174				
Db	126	PAAYITELVREAKDLHPDLSLFRD	LIRPDLAALSQNMDDDELSTLS	LSNELLYRGIG	185				
Qy	175	RKTGGSDALMESLTYRQAIPTYHOP	YETIRQIVIMTHDSTLSALSRN	PEVVMQAGAS	234				
Db	186	AASGLDDDSVRELLAGVRLTGLTPY	HWAYEAAQAILVQDPTLMGFSR	NPDPVAQLMDPAS	245				
Qy	235	LLAILANISPELNYILTEBITEKNA	DALPAQNSENITPENFASQSWIA	KYVYGLSELSEVQ	294				
Db	246	MLAIEADISPELYQIIAAEETITDS	YEALMSKFNFG-MPPSSLSYDA	LATFYDLDYDEL	304				
Qy	295	KYLGMLQNGYSTSAVDNISGLVNN	ESK--LEAYKITRVKTDYDKN	INVFDLWME	352				
Db	305	SLLS-LRLDPSNPNNXYINSQLSV	TNLNESTGLTIHHYLRITLGD	-SQINP-ELIPY	361				
Qy	353	GNNQFFIRANFKVSRFBGATLRNA	QPSGIVGS-----LSGPLYANT	FNKSNYLSNIDS	407				
Db	362	GDGTLYL--NFSVV-----STI	SEDSFKJGSLGSSNLSYSGD	YQLQGVYSPIVEIDEG	415				
Qy	408	EYKNGVKIYAIRVYTSSTATNQGGI	-----FTFESYPLTI	PAIKLNKATRLCLTSGLSP	462				
Db	416	KLMDGITI-----GLSRKGGGY	SVTNFTLIEYDPAIFIKLNK	IVRLRYKATGMTT	466				
Qy	463	NELQTVRSDNAOGIINDSVLTKV	FYTLFYSHRYALSFDDAQVL	NGSVINOYADDDSVSH	522				
Db	467	ABYQITNIIINGLTIDHAVLSKIF	LVRYLMRHYQLDVARSLILCNG	TIISDQAFSGBTGL	526				
Qy	523	FNRLFTNTPPLKGIIFEADGNVT	SIDPDEQSTFARSALMRGLGV	NSGELYQLQGLAGVL	582				
Db	527	FTTLFTNTPPLNGQLFSADDTPL	RSEAPEDAFLSVLKEAFNIS	SAGSLTLQLAGS-D	585				
Qy	583	AQNTITLSVVFVSSLYRLTLRLAR	HVQLTVNELCWLGLSPNKT	KTATSLSGBELPLRVIV	642				
Db	586	SSAGFCSADNIAALYRVKLLADH	DLSDAGELSMLLSVSPFSG	VAAAGSLSDNELTQ--F	642				
Qy	643	LYQVTOWLTEABITTEAIWLLCT	PESGNSIPISNLLANNLRPSI	SEDMAQSHNRELQAE	702				
Db	643	LYQTITLWTEQGWTSVDFVLM	LTQYGTGLTTPDIEILLAS	RNGLS-----GRELFPE	695				
Qy	703	IL----APFTAALHLASPDMA	RYILLTWTNLRPGGLDIAG	FMVTLVKLSINANETQLV	758				
Db	696	TLSEGDGAPFIAAANQQLD	DTATKAMLTWADQAPGLT	TEFILLVWNAAPNDEQACQMA	755				
Qy	759	QFCHVMAQLSLSVQTLRLS	EAELSVLVISGFVLGAKNQ	PAG-----QHNIDTLFSL	810				
Db	756	GFCQALWQALAIIRSTGLS	TRETLIV-----SQGRFRT	GWHHLPHDLPALRDI	805				
Qy	811	YRFHWINGLNPGSDTLDMLR	QOQTLTADRKLASVMGLD	ISMTQAMV-----SAGVQ	--L	864			
Db	806	TRFHAVNRSGSHAGEVLT	LAETGELSSALLARALSQ	NEQDVTGALAQVRGAGEQ	DNVSF	865			
Qy	865	QCWODINTVLQWIDVASA	HTMPSVIRTVNIYVTLNKA	ESNLPSDWSQTLAENMEA	924				
Db	866	TSWEEDVAEQWLDNGETS	ITPSGLASLTALKY---N	VSDSGAPLYSQWQVVSGLLQA	922				
Qy	925	GLSTQQAQTLIADYATBRLS	VLNCWFLANTIQPSGVLS	HRSDDLTSYFLIDNQVSSA	IKTT	984			
Db	923	GLAKSQSSALHDYLEGTS	SALCAYILRNLPNVS--	GRDDLFGYLLLDNQVSAK	VKTT	980			
Qy	985	RLAEATAGIQLYINRALNR	IEPNARADVSTROQFTDW	-TVNNRYSTWGGVSRVL	YTPENY	1043			
Db	981	RIAEATAGIRLYINRALNG	IELSAMAEVRQOFTDWD	TFNKRYSTWAGVSELV	YTPENY	1040			

Best Local Similarity 40.1%; Pred. No. 1 le-234;

Query Match 35.1%; Score 4583.5; DB 2; Length 2376;  
Best Local Similarity 40.1%; Pred. No. 1.1e-234;

```
Qy 1044 IDPTQRIGOTRMMDELLENISQSLSDTVEDAKFTYLTRETVADLKVVSAVHDVNSN 1103
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1041 LDPTVRIGQTMMDTLQSVSQSSNRDVTDEDAFKYLTTPTEQIANLUNTVSGYHDNASMT 1100
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1104 TGLTWFGQTRENIPEYYWRNVDSRMQAGLAANAKWKTKIDTAVNPYKDAIRPVIFR 1163
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1101 QGTTWYVGRSITDQTNWYRANSHSKIODSMWPNANWTGWTKINGMWPNSDLVCSVFFN 1160
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1164 BRHLIHWKEVAKNGDPVETVDR-PTLKLAFLRHGDSWASAPWSVDITTVQVAVTDKK 1222
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1161 SRLVWVVEENQSDATAEASTTTTQQSYTLKLSFRYDGTWSSPSVFDITGNIAF----- 1215
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1223 PDTERLALAAAGFQGEDTLLVYVYKTKGYSYDFG-----SNKNVAGMTIYGD-GSEKKME 1277
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1216 PETQGMHVCNPLT-EQLYCAFYVTSK--PDFDNAQLISVDNDMTLNVISDIDGIFKSVS 1272
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1278 ---NTALSRYQLKNTF-----DIHTQGNDLVRKASYPFAQDFEVPAS 1318
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1273 HEFNTSTEKF--INNVSFSDPSANYFVSATSLLDDVIHS-----DFSL--- 1312
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1319 LMGSAIGDDSLTWENGINIPOITSKYSSDNLATILHNAATVTVYDGSNVIRNKQISAM 1378
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1313 LN-----SKTSTVFTNEDSSLLTPE-----LHITANVSCFVSTAGIATQSTIEKF 1358
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1379 KLTGVD-----GKSQYG-NAFI--IANTVKHGYSYDLGGPITVYNTKKNY-IASV 1425
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1359 VOAGIEFBEINFYAQAAGGDFGVVDVNSKVTQVGKEAVG--VTV--KSYVTGV 1412
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1426 QGHLMNADYTRRLILTPVNNYIARLFEFPFSPNTILMTVTVGSNKTSDPKKCSYAVDG 1485
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1413 SGSVE-----LFIDSSNKYFSGIL-----SDKMITALISGST-----SKVNY-VSS 1452
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1486 NNSQGFQIFSS-----YQSSGMLDDITGNNTDIIKIVNAGSKTHFTFASDHIALSPA 1538
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1453 IGSQDFSVKSLMPALQIYELIDDIILTSVGNVTEIK-----SWPSAEWYNDKLSLQSG 1506
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1539 NSFDPMPYTFKPLEIDASSLAFTNNAIPL-----DIVPETKAKDGR-VLGKIKQTLSVK 1591
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1507 NNL-----FNTKSLSTVNTSDIVDEDFVTFTFVAVDQNNVVLAAFTAILTVI 1555
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1592 RVNYPNEDILFLRETHSQAQWQL-----GVYRIRLNTLLASQLSRANTGIDTILTMETQ 1647
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1556 RNINNTSVIALRKNTRCAQVIRFAGNDVALIRLNTLFAQLVDRANTGIDTILSMETQ 1615
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1648 RLPEPPLGEGFPANFVLPKYDPAEHRDWPFKIHI GNVGNTGRQPYSGMLSTSETSM 1707
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1616 RLTEPALFEEG-----SDVF----- 1629
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1708 TLFVPYABGYMHGVRILGVGQKITDNTWESAFFYFDETKQFVLINDADHDSGMTQQ 1767
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1630 ----- 1629
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1768 GIVRNKKYKGLNVIATGYTSAPMDFNSALYFELFYTPMFCQRLQEQFDEAT 1827
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1630 -----MDPSGANALYFELFYTPMFCQRLQEQFDEAT 1665
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1828 QWIVNVPAGYIYNGETAPHWCNRPLEETTSWNPANPLDAIDPAVANDPMYKATTF 1887
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1666 RWLQYVNPAGVNVGLQNTWNVRLPEEDTWNDSPLSDIDPDATAQYDPMYKATTF 1725
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1888 MRLDOLLRGDMAYRELTRDALNEAKWYVYRLELLGDERPEYDGSQWQAAPSLSGAASQ 1947
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1726 MSYDLILLTARGDAAYRLERDTLNEARWYVQALNLLJGDEYIISFDADWSALTIGDASE 1785
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1948 TVQAAQQDLTMLGRGGVSKNLRANSVLGFLPEYNPALTDYQTLRLRLNLRHNLISI 2007
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1786 VTRRDYQEAALLAVRLVPAPEPTRANSITLFLPQNEVLKGYQTLQAQLHNRHNLISI 1845
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2008 DGQPLSLAIYAEPTDPKALLTSMVQASQGGSAVLPGLTSLRFPVWLERNLVAQLTQF 2067
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1846 DGQPLSLSVYATPSEPSLQSAVNSQAQAALPAAVMPLYSFPPVMLENARGWVSLTGF 1905
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
Qy 2068 GTSILSMAEHDDADLTLLLLQQGMELATQSIHQRTQTVDEVDADIAVLAESSRSQAQNL 2127
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1906 GNTLLGITERQDAEALAKLLQTQGSSELIHQGLRQODNVLEIDADIAALESRRGAQMRP 1965
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2128 EKYQOOLYDEINHGEOBAMSLLDAAAGOSLAGOVLSIARGVADLVNPNVFLGACGSRWGA 2187
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1966 ERYKVLYEADVNTGEKQAMDLYLSSSVLSASTAALFLAEAAADMLPNYIYGLAVGGSRYGA 2025
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2188 ALRASAVMSLSATASQYSADKISRSEAYRRRQEWIORDNADGEVKQMDAQLESKIR 2247
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2026 LFKATATIGIQVSSDATHISADKISQSEVYRRRREWEIORDSAQSDVAQIDAQAARAVR 2085
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2248 REAAQMVQEYQETQQAHTQAQLELLQKFTNKALYSWMRGKLSAIYYQFFDLTOSFCIMA 2307
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2086 REGAELQKTYLETQTOQAQLAFLQSKFNNTALYSWLRGRLSAIYYQFYVDLAVSRCLMA 2145
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2308 QEALURRELDTONGVTPIRGAWNGTTAGIMAGETLLLNLAEMEKVWLERDERALEVTRTVS 2367
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2146 QQAQWQDKFETR-SFIQFGAWMGANAGLAGETLMLNLAQMEQAWLTGDERAIEVTRVC 2204
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2368 LAQFYQALSSD-NFNLTETKLTQFLREGKGNVGASGNELKLSNRQIEASVRLSDLKIFSDY 2426
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2205 LSEVYTSIAEDAAEASLADKVVVELVNSGSGAGTKSNGLQMDQQOLEATILKLADIGNDY 2264
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2427 PESLGNTRQLKQSVTLTPALVGPYEDIRAVLNLYGGSIWMPRGCSAIALSHGVNDSGQFML 2486
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2265 PVSIGTWRRIKQISVTLTPALVGPYQDVRAVLISYSGSMVMPRGCSALAVSHGMNDSGQFOL 2324
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2487 DFNDPRLPPEGLPVDVTGTLTSPFPDADQKALLESLSDIILHIRYTTIS 2538
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 2325 DFNDPRLPPEGLPVDVTGTLTSPFPDADQKALLESLSDIILHIRYTTIS 2376
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 10
Q8GFAO PHOLU
ID Q8GFAO PHOLU PRELIMINARY; PRT; 2499 AA.
AC Q8GFAO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE TcdA2.
GN Name=tcdA2;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=23488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Daborn P.J., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AA017201.1; -; Genomic_DNA.
DR InterPro; IPR000577; FGSY_kin.
DR InterPro; IPR001063; Ribosomal_L22.
DR PROSITE; PS00445; FGSY_KINASES_2; UNKNOWN 1.
DR PROSITE; PS00464; RIBOSOMAL_L22; UNKNOWN 1.
SQ SEQUENCE 2499 AA; 283745 MW; E7C4C9201464B4D8 CRC64;

Query Match 35.1%; Score 4572; DB 2; Length 2499;
Match Local Similarity 38.7%; Pred. No. 4.8e-234;
Matches 1063; Conservative 444; Mismatches 740; Indels 502; Gaps 72;

Qy 22 LADLQYLSFSELRKIFDDQLSWGBARHLYHETIEBQKKNRLLLEARIFTRANPQLSGAIRL 81
```

Db 19 L7DICHLSFDELKXVKDKLSWSQSLYLAQVQKDKNLLHRAILKRANPHLQSAVHL 78  
Qy 82 GIERDSV-SRSVDENFGSRSFVKPQSVASFPSPAGVLTLYREAKDLHPSSSAYHLDN 140  
Db 79 ALTAHADOQGVNSRFGNRASKYAAPGAISFSLAAYLTLYROARNLHAEGSYHLDT 138  
Qy 141 RRPDLADLTLSQNNMDTBISTLTLSNELLHEITKTKGSDSA-LWESLSLTYRQADITPY 199  
Db 139 RRPDLKSLVLSQNNMTBISTLSNMLNLSIKTQPNLNSHAKVMEKLSLFTSGSMYP 198  
Qy 200 HOPYETIRQVIMTHDSTL---SALGRNP-EVWGAEGASLLAILANSPELYNLTBEI 254  
Db 199 HDAYESVRKIIQLOQAPVEFQSSLTLDTPITKLMYQ---ISLGINASVSPLEFILTQKI 255  
Qy 255 -TEKNAD---ALFAQNFENITPENPASQSWAKYVGLSELSVQVGLQMLQNGYSDS 307  
Db 256 KPATNADNTELKUYKKNFGE-IKSIQWABAYLKSYNLT----- 296  
Qy 308 TSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINYFOLMYBGNNOFFIRANFKVSR 367  
Db 297 -----DKELNOF----- 303  
Qy 368 EFGATLRKNAQPSGIVGSLGPIANTNPKNLSNISDSSEYKNGVKIYAYRTSSSAT 427  
Db 304 -----SKIKQI-DLWNIIGDEITQYH----- 324  
Qy 428 NOGGIPTFESVPLTIFALKLNKATRLCTISGLSPNELQTVRSDNAOGIINDSVLTQVF 487  
Db 325 -----LLKFNKAINLSRSTELSP-----IILNSAIDIL 353  
Qy 488 YTLFYSHRYALSFDQAQVNGSVINOYADD-DSVSHFNRLFNTPPLKGIPEADGNTVSI 546  
Db 354 -----KKTPEDDSD-----NPRDDPDYLSFQDL-----DLSD 383  
Qy 547 DPDEQSTFARSAL-----MRGLVNSGELYQLGKLAGVLDAQNTITLSVVIS----- 595  
Db 384 EPDIDEDVL-REALRVKQMYRGI-----DAETALILCKAPISENPSHP 427  
Qy 596 SLYRLTLARVHOLTVELNCLMYGLSPNGKTTAS-LSSGELPRVLVWLYOVOTWLTAE 654  
Db 428 DLSC--LHADTHQLTIDELGVL--VAIDEGKTDLSQITHNLAVLISKLYSVNWLRTK 484  
Qy 655 ITTEAIWLLCTPEFGSNTSPEISNLNML-----RPSISED-- 690  
Db 485 WSVYQLFVMTDKYKNTLTPELNNLLDVTYNGLQNFYKONLLKIKONLLKAKESLPEDKO 544  
Qy 691 -----MAQSHNRELOA----- 701  
Db 545 NLPKAEQYLLBAEKYLLAAEKYLLAANKNPLEAKKALKEYEKNOEAYEKNLKE 604  
Qy 702 -----ELAPPIAATLHASPDMARVILLWTDNLRPG- 733  
Db 605 HEKYLKAGENLPAIKENLLKIKENLPKAIISPYIAAALQPSNVALSWAWADKLSGK 664  
Qy 734 --GLDIAGFWTLVLK-----ESLNANETT-----OLVQPCHYMA 765  
Db 665 ENKMTADSFNNWLRKKPIETQSKTTEATEATEATEATEATEATEATEATEATEATEATE 724  
Qy 766 QLSLSVQTLRLSEABLSVLVSGFVAVLGAKNQAPAGOHNIDTLFSLYRPHOWINGLNGPS 825  
Db 725 QALAIYRSTGLSESTLR-LFVTNPQIFGLTAKTTSTHNVLSLIMLTRFTDWNVSLGENAS 783  
Qy 826 DTLMLROQTLTADRLASVGLDISWVTOAMVSAGVNOLOQWODINTVLOHIDVASALHT 885  
Db 784 SVLTBEFGTGLTAEALANMNDKNLLEQASTQAOAD-FSNWPSIDNLOWINISRQLNI 842  
Qy 886 MPSVTRTVNRYVTALNKAENLPSWDEWOTLAEWAGLSTQAOQLADYATRLSSV 945  
Db 843 SPQGVSELAKILDI-----ESS-TNYAQWENVASILTAGLDTQANTLHAFELGESSTA 895  
Qy 946 LCNWFLAN-----TQPEGVLSHSDLDLYSYFELDNQVSSAIKTRRLAEIAGIOLYINR 999

Db 896 LSTYYIYSHNQDKREERKHTVIKORDDLYQVLLDNQVSAAKTEITAEIASIOLYINR 955  
Qy 1000 ALNRIEPNARADVSTROFFTDW-TVNNRYSTWGGVSRVLYYPENYIDTQORIGOTRMDE 1058  
Db 956 ALKNWEGDTSVTSRSLFFTNWCKNKYKSTWAGITKLKLYYPENYIDTQORIGOTRMDE 1015  
Qy 1059 LLENISQKSLRDTVEDAFKTYLRTFETVADLVKVSAYHDNNSVNTGLTWFGOTRENLP 1118  
Db 1016 LLOSTISQSLNTDTVEDAFKSYLTSFEQVANLEVISAYHDNINNDQGLTYIGRSKTEVN 1075  
Qy 1119 EYVWRNYDISRMOAGELAANAWEKWKIDTAVNPYKDAIRPVIPRERLHLWVEKEEVA- 1177  
Db 1076 QYWRSDHNNKPSGKPPANAWSEWHKIDCPINPEDIPTPVQVQSRUYIILWLEOKKVTN 1135  
Qy 1178 -----KNGTDPVETVDRTKLAFLRHDSGWSAPWSYDITTOVEAYDTKPKPTERLAL 1230  
Db 1136 RAEGEAIKQSGKTTTSY-HYELKLAHRYDGTWNTPTITFDVDEKISGLNLEL--NKALGL 1192  
Qy 1231 AASFGQGEDTLLVFFYTKGKSYDFGSGNKN-----VAGMTIYDGGSPKQ--ENTALS 1283  
Db 1193 YCASYQGGKDLVLMFYKKQEQLNNYTEKTYTAPIKGLYITSNMSPEEMTPESYRLNA 1252  
Qy 1284 YSQLKNTPFDIITHTOGNDLVRKASVRFADQFVFPASLANMGSAL--GDDSLTYMENGNIPQI 1341  
Db 1253 HKQ-----PDT-----NNVVR-VNNRYAESYEPSVSNNGVDMGEGYLSWYGGSI--L 1300  
Qy 1342 TSKYSSDNLAJ-----TLHNAFTVRYDGSNNVIRNKQISAMKLTGVDGSKSQYGNAPI 1394  
Db 1301 ITRPDSNKSQIOPKLIIRHNG-----YEGR-----QBNQCNLMKKYG-----KLGRFI 1346  
Qy 1395 IANTVKHYGYSIDLG-GEITYVNTKQNY-IASVQGHLMNADYTRRLIITPVENNYARLUF 1452  
Db 1347 I-----YTLGINPNLSNKKLIYPVQYEGNESKLSQGRLLFYRSTTNTRAWF 1397  
Qy 1453 EFPSPNTILNTVTVGSKNTSDFKCKSYAVDGN-----NSQCFQIESSYQS 1499  
Db 1398 -----PNLSSDSKEMSIITGGNISNGYINDKHSNKKPEEFYFMD 1440  
Qy 1500 SGWLD-----IDTGINNTDIKI---TVMAGSK--THFTASDHASLSPANSFDPAMY 1546  
Db 1441 HGGIDTVSEBIFINTKLOPSNVKLIIVKTVKDDGKLDKSKPYIAEDKVSVPKPTNPEEMCY 1500  
Qy 1547 TKFPL-ETDASLAFPTNNAIPLDIVETKA-----KQGRVLGKIKQTLIS-VKRVNYPPE 1598  
Db 1501 QFNLDQIDVSTLVFKNEASIDITFTASADAFESGKEQRLNGBEHFSIRIKKANVN-- 1558  
Qy 1599 DILFLARETHSCAQMOLGVYRILNTLLASOLVSRANTGIDTILTMETQRLPERPLGEGF 1658  
Db 1559 DVLTLHDDPSGAQYMQMGAYRTRNLTPARKLISGRANAGIDITILSMETQNTQBPOLGKGF 1618  
Qy 1659 PANFVLPKYDPAEHDGDERWFKIHIHNGVNGTGROPYSGMLSDTSETSMTLFVPPYAEY 1718  
Db 1619 YVNTFLPKYDQNTGNERQFKIHIHNIAGDNTMRPYQOILADT-ETSVLFPVPEKQSY 1677  
Qy 1719 MHGVRGVGVQKITYDNTWESAFYPDETQKQFVLINDADHDSGWTQOIGIVNKKYKG 1778  
Db 1678 TNEGVRGVGVKYSYLVGWEPAPFYEIQQKFTILINDADHNSANTQSGEKTGKTKYKG 1737  
Qy 1779 FLNYSIATG--YSAPMDENSALYYMELFYVTPMCMCFORLLOEQKQDEATOWINVYNP 1836  
Db 1738 FLDYSILIDHQHTEPFMDFNANGANSYFWELFYVTPMLIAQRLHHEQNFDEANRLKYVNP 1797  
Qy 1837 AGYVNGEIA-PWTWNCRPLEBETTSWNPANPLDAIDPDVAQNDPMHYKIATFMRLLDQII 1895  
Db 1798 SGHTANGOKPHNNVVRPLQEDTSWDDPLDTPDPAIAQHDPMHYKVATFMCALDILL 1857  
Qy 1896 LRGMAYRELTRDALNEAKOMMYVTRLELLGDEPEDYGSQWAAAPSLSGAASQTVQAAQ 1955  
Db 1858 EQGDYAVRQLERDRLAEAKWYMQALHLLGDKPHLLLSSTWSDBELKEAADLEKQAAHAK 1917  
Qy 1956 DTLMLRGGVSKNLRANSVGLFLPEYNPALTDYQWTLRLRLRNLNLSIDGQPLSIA 2015  
Db 1918 AIADL-ROGQPKD-----GSNTDLFLPQVNEVWLSYQWLEQRLNRLNLSIDGQPLHLP 1972



Db 890 IAEASTQLYINRALKNMEEDTDSVTSRSTFTNDWKYKRYSTWASIAKLLYPENVI 949  
QY 1045 DPTORIGOTRMDELLENISQKSRDVEDAFKTYLRFETVADLKVVVSAYHDVNSNT 1104  
Db 950 EPTLRIQTKMDALLQSISQSLQNTDVEDAFKSYLTSFEQVANLEVISYHDNINNDQ 1009  
QY 1105 GLUTVFWQOTRENLPYVWNVDSIRMOAGELANAWKWKIDTAVNPYKDAIRVPIRE 1164  
Db 1010 GLTYPIGRSKTEVQVYKRSVDHNFSEKGFANAWSEWHKIDCPINPYEDTIRVIFQS 1069  
QY 1165 RLHLTWVKEEVA-----KNGTOPVETVDRFTLKLAFLEHDGWS 1204  
Db 1070 RUYITWLEOKVTHRSEGETLPKVALSEIKTDKQSKTTTSY-YIELKLARIIDATWN 1128  
QY 1205 APWSYDITITQVBAVDKPKDTER-LALASGFGQEDTLIVFYKTKGYSDFGSGKNVA 1263  
Db 1129 TPIITFDVD--KKISDLNLEANKVLGLYCASYQGRDKLLVMFYQKPEQASY--INAPMQ 1183  
QY 1264 GMTIYDGSFKOM--ENTALSRYSQLKNFTDIIHQGNLDLVRKASYRFAQDFEVPASLNM 1321  
Db 1184 GLYISSNMSQEBMNPNDYKLNTHKQIDT-----HTATNSVIR-VNNRYAESYEIPSSVNN 1237  
QY 1322 GSAI--GDSLTVMENGIPOITSKYSSDNLAI-----TLHNAFTVRYDGSNGVLRN 1372  
Db 1238 NNGYDWBEGYLSMWYVGGSI--LITRNQSDNSKIQISPKLRIIHNG-----YEGR-----QR 1286  
QY 1373 KQISAMKLTGVGDKSQYG---NAFTIANTVHYGYSDIGGPITVYN-KTRNYIASVOGH 1428  
Db 1287 NOCNLMK-----KYGPKDKFII-----YTLG--INQNLSNKKFIYPVQY 1327  
QY 1429 LMAADYTR--RLILFPVNNYVARLFEPPSPNTILNFTVVGSKNTSDFKKCSYAVDGN 1486  
Db 1328 EGNASQSQRLLFYRDSATNFVRAW-LPNLPSGQEMSITVGGNISGNYGVIDNKHNN 1386  
QY 1487 NSQGFQIPSSY-----QSSGLDIDGGINNTDKITVMAGSK-----THFTTAS 1530  
Db 1387 NP-----FKGYFYMDHGGIDFDVSGPMPINTNIQPSNVKIITVTKVEDGKLTSEPYTAN 1441  
QY 1531 DHIASLPANSFAMPYTFKPL-EIDASSIAFTNFIAPIDIVPETRAKDGRLVGKIKQTLIS 1589  
Db 1442 DKVSVKPLNFEMCYQFNGLEIDVSKLVFNKNEASIDITP-TAFADAFSGQKERNLG 1500  
QY 1590 -----VKRVYNPEDILFURETHSGAQYMWOLGVYIRLNTLTLAASQLVSPANTGIDTI 1641  
Db 1501 EBHFSIRIUKKSETN--NVLSLHNSGSAQYMWQVAYRTRNLTPARKLISRANAGIDTI 1558  
QY 1642 LTMETORLEPPGLGFPFANFVLPKYDPAHEGDEWFKHIGVCGNTRQRPYSGMLSD 1701  
Db 1559 LMETONTQELQLGFGFATFVIPPYNPSTHGDWRWFKLYIKHVSDDNS-HIYSGQLKD 1617  
QY 1702 TSETSMTLFVP-----YAEGYMHGVRGLGVGYOKITYDNTWBSAFYFDETQKQ 1751  
Db 1618 TN-IGITLFIPLDDIPLNQDFHAKYV-----MTFKQSPDSGWSGPHFIRD--NKG 1665  
QY 1752 FVLINDADHSGMTQOGIVKNIKKYKFLNVSIAFGYAPMDFNASALYIYWEYFPTPM 1811  
Db 1666 IITINSS-----SILTHPESVNVNLNVG-----SEPMDFNGANSLSYFWELEFYFPTM 1711  
QY 1812 MCFQRLLOEQKQDEATOWINYVNPAGYVNGEIA-PWITNCRPLEETTSWNAFLDAID 1870  
Db 1712 LIAQRLLEQNFQDEANRLKYYVNFSGHANGOKQYPHTWVVRPLOEDTSMNDDPLGSPD 1771  
QY 1871 PDVAQNPDPMHYKIAFTFRLLDQLILRGDMARELTDALNEAKWMYVTRTLELLGDPEPD 1930  
Db 1772 PDIAIAYDPMHYKVASFRTLDLLEQGYAYRQLERTDLAEAKWMYQALHLLGDKPHL 1831  
QY 1931 YGSQWQAAPSLGSAASQTVQAAVQDQLTMLGRGVSKNLRTRANSVLGLFLPEYNPALTDY 1990  
Db 1832 SLSTSWNDPELEBAADLEKQTHAKAVADLRQ-----QSTGGSNTDLFLPQVNWMLSY 1886  
QY 1991 WOTLRLFLNLHNLNIDQPLSLAIYABPTDPKALLTSMWQASOGGSNVLPGTSLRFP 2050  
Db 1887 WQKLEQRLYNLHNLNIDQPLSLAIYABPTDPKALLTSMWQASOGGSNVLPGTSLRFP 1946

QY 2051 PVMLETRNLVAQLTQFGTSLLSMAEHDDADBELTTLLLOQGMELATQSIQIRIQORTVDEVD 2110  
Db 1947 PHMLENARSVMYSQLTQFGSTLQNIITERRQDAEALNTLLOQAELVLTNLISIQDKTIEED 2006  
QY 2111 ADIAVLASRRSAQNLKQYQOYDEDINHGEORAWSLDDAAAGOSLAGOVLSIAGVAD 2170  
Db 2007 VEKTVLEKTRAGAQRFSDYSKFYNEDINAGEKQAWLRASVAGISTALQASHLAGAALD 2066  
QY 2171 LVPNVFGLACGSRGAALRASASVMSLSATASQYSADKISRSEAYRRRRQSEWIORDNA 2230  
Db 2067 LAPNIFGADGGSHWGAIAQATGNVWFPSASVMNTEADKISQSEAYRRRRQSEWIORDNA 2126  
QY 2231 DGEVKMDAQLSLEKIRREAAQMVQYQETQQAHTQAQLELLQRFPTNKALYSWNRGKLS 2290  
Db 2127 ETELKQIDAQLGSLVVRREAAVLQKTSKLTQEQHTQAQTLFQHKFSNQALYNWLRGRLS 2186  
QY 2291 AIYQYFPDLTQSFCLMAQOEALRRELTDNGVTFIRGAWNGTTAGLMAGETLLNLAEEMK 2350  
Db 2187 AIYQYFQDLYTVARCLMAEMAYRWETNDAAARIPKFGAWGTHAGLAGETLLNLQAQMD 2246  
QY 2351 VWLEDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVSGASGNELKL---- 2406  
Db 2247 AHLKQEQVLEVERTVSLANIYK--EKQFSLTEKIAELMKNKLVTAGSGNNTLKFGTGN 2304  
QY 2407 SNRQTEASVRLSDLKIFSDYPES--LGNTRQLKQVSVTLPALVGYEDIRAVLNYGGSIV 2464  
Db 2305 AQTSLQASISLADLQIRHDYPENSGVGNVRIKQISVTLPALVGYEDIRAVLNYGGSIV 2364  
QY 2465 -MPCGSAIALSHGVNDSGQFMDLDFNSRYLPFEGISVNDGSLTSLSPDA-----TDR 2517  
Db 2365 GLAESCKLSVSHGNDGQFQDFNDGKFLPFEGIDI-DKGTLLTSLFPNALGEGSKGK 2423  
QY 2518 OKALLESLSIDILHIRYTIIR 2537  
Db 2424 QXTMLESLSIDILHIRYTIIR 2443

## RESULT 12

Q7N7Y0 PHOLL PRELIMINARY; PRT: 2215 AA.  
AC Q7N7Y0  
DT 01-MAR-2004 (Tremblrel. 26, Created)  
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE Insecticidal toxin complex protein TcdA5.  
GN Name=tcdA5; OrderedLocName=plu0971;  
OS Photobacterium luminescens (subsp. laumondii);  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photobacterius.  
OX NCBI TaxID=141679;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN-TT01;  
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Rusnlok C., Frangeul L., Buchrieser C., Gavaudan A.,  
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,  
RT "The genome sequence of the entomopathogenic bacterium Photobacterium  
luminescens";  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571862; CAE13266.1; -; Genomic\_DNA.  
DR PhotoList; plu0971; -.  
KW Complete proteome.  
SQ SEQUENCE 2215 AA; 252782 MW; 188CEB6A61E69956 CRC64;  
Query Match 34.4%; Score 4491; DB 2; Length 2215;  
Best Local Similarity 39.8%; Pred. No. 8.2e-230;  
Matches 1012; Conservative 394; Mismatches 737; Indels 402; Gaps 47;  
QY 53 TIEQKNNRLLEARIFTTANPOLSGAIRLIGRDSV-SRSYDEMFGARSSSVFKPGSVAS 111







Db 1245 ---NTVLVSVDSDMKLNTVSDNLI--PQSIKHEFN-----TSTERFINNFSD 1287  
Qy 1317 ASLWNGSAIGDDSLTWEMNGNIPQITSKYSS-----DNLAITLHNAAFVTVRVDGSGNVIRN 1372  
Db 1288 SSASYAIS-GTDSIDDATHSDILLNGRVMTVITDN-SMGSWNPENFIAASVCFITNS 1345  
Qy 1373 KQISAMKLTGVDGSKSYQGNAFIIANTVKHYGGYSDLGPGPIVYVYKNTKNYIASV---QGH 1428  
Db 1346 SLVS-----DEYNYFI-----MGG---VFPERKNFYVAIWDRPGG 1379  
Qy 1429 LWNADYTRRLILTPVENNYIARLEFPFPSPNTILMTFTVGSNKTSDPK-KCSYAVDGN 1487  
Db 1380 FIGVD-----VSN-----SKIYQVG--KASDFVIHKTFVSQVQ 1411  
Qy 1488 SQ-----GFOIFSSYQS-----SGWLDIDTGINNTDIKITYWAGSKTHFTASD- 1531  
Db 1412 GQIELNVHSGKHLFGTLTDAMVAALLNGSTSTTSILN-GATITASFGAVLPVYNANDL 1470  
Qy 1532 -----HIASLPA-----NSPDAMPYTPKPLEIDASSLAFTNNI 1564  
Db 1471 IDDTLTFARMNGTEIKMPSSEWYNRELNLQVGNLFTNALSF---TVDTSOI--TDDE 1525  
Qy 1565 APLDIVETKAKGRVLGKIQTLTSVKRVNYPEDILFLRETHSGAQMQL-----GYVRI 1620  
Db 1526 FDTVLTFTAVGENDAVLAARTVINVRRLINNDPVIARLKNTRGAQYIRFTAGNDVALI 1585  
Qy 1621 RLNTLLASQLVSRANTGIDTILTMETORLPPEPPLEGEPFANFVLPKYDPAEHBGDERWPKI 1680  
Db 1586 RLNTLFARQLVDRANTGIDTILSMETORLTPEALEEG----- 1622  
Qy 1681 HIGNVGNGTQRPYYSGMLSDTSETMTLFPVYAEYGYMHSGVRLGVGYQIKITYDNTWES 1740  
Db 1623 -----SDVF----- 1626  
Qy 1741 AFFYFDETKQOQVFLINDADHDSGMTQOQIVKNIKKYKGLNVSATGYSPAMPDFNSASAL 1800  
Db 1627 -----MDFSGANAL 1635  
Qy 1801 YWBELFYTPMFCORLLOEQFBQATQWYNVYNPAGYVINGBIAPWNCRLPETHTS 1860  
Db 1636 YFWELFYTPMFWFQRLLEQNFPEATRWLYVWNPAGHVGNLQNYTNWVNRPLEEDTS 1695  
Qy 1661 WNANPLDAIDPDAVQNDPMHYKATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1920  
Db 1696 WNSPLDSVDPAVAQNEPMHYKATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1755  
Qy 1921 LELLGDEPEDYGSQQAAPSLSGAASQTVQAAQQDLTMLGRGVGSKNRLTANSVLGLFL 1980  
Db 1756 LNLGGERPYISFDADWSALTIGDAASEVTARSYQSALLAVRRQQGAPDVRTANSITLTL 1815  
Qy 1981 PEYNPALTDYQWTLRLRLFNRLNLSIDGQPLSLAIYABPTDPKALLTSWQASQGGSAV 2040  
Db 1816 PQQNAVLLKGYWQTLAQRLLYNLRLNLSIDGQPLSLSVVATPAEPTALLNAVSSAQGSSSL 1875  
Qy 2041 LPGTLISLYRFPVMLERTNLVAQLTQFGTSLLSWAHDHDADELATLLQLQWELATQSIR 2100  
Db 1876 PVAVNPLRFPVMLERNARGQVSLIQQFGSTLISLITERQDAELALAEILLQTSSELLOGLR 1935  
Qy 2101 IQORTVDEVDADIIVLAESRRSAQNRLKYOQLYDEDDINHGEORAMSILDAAGQSLAGQ 2160  
Db 1936 QQDNALAEIDADILAEESRGAQARFEYYSRLYDADVNTKEQAMDLYLSSSVLSASSQ 1995  
Qy 2161 VLSTABGVADLVNPFGLACGSRWGAALRASAVMSLSATASQYSDAKTISRSEAYRRRR 2220  
Db 1996 VLFMAGAADMPLNPIYGLAVGSRYGALFNATAIGIQVSSDATRISADKISQSEMYRRRR 2055  
Qy 2221 QEWIQRDNAGEYKQMDAQLESKIRREAAQMOVEQETOQAQTQALELQKFTNKA 2280  
Db 2056 EDWEIQRDGAOSDVTGMDAQAVNVRREGAELOKTYMEMQOIQOQAQVLYQKFSNTA 2115  
Qy 2281 LYSWMRGKLSAIYYQFDFLTQSFCLMAQEARLRELTDNGVTFIRGGAWNGTTAGLMAGET 2340

Db 2116 LYSWLRGKLAALYYQFYDLTASRCLMAQSAQYQWD-RNSTTRFIQFGAWQGTYAGLAGET 2174  
Qy 2341 LLLNLASWEKVLBERDERALEVETVTSIAQFYQALSSD-NENLTKELTQFLREGKGVGA 2399  
Db 2175 LMLGLSRMEQAWLESDBREREVTRVCLSEVYIAGLAGDAAEVLADEVVGLVNGGTSSAGT 2234  
Qy 2400 SGNELKLSNRQIEASVRLSDLKIFSDYPESIGNTRQLKQSVTLTPALVGPYEDIRAVLNY 2459  
Db 2235 ATNGLKPADQOQLATLSLADLNIRSDYPSLGFTRIKQISVTLTPALVGPYQDYRAMLSY 2294  
Qy 2460 GGSIVMPRGSAIALSHGVNDSGQFMDDFNDSRVLPFEGISVNDSGSLTSLSPDATDRQK 2519  
Db 2295 GGSVMVPRGCNALAVSHGMNDSGQFQDFNDSRWLPFEGIPVGDSTLTLSPDADGRQ 2354  
Qy 2520 ALLLESLSDIILHRYTIRS 2538  
Db 2355 AMLLSLSDIILHRYTIRS 2373

RESULT 14  
Q4ZNN7.PSES  
ID Q4ZNN7.PSESY PRELIMINARY; PRT: 2502 AA.  
AC Q4ZNN7;  
DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Insecticidal toxin complex protein Tcdal.  
GN ORFNames=PsyT\_4205;  
OS Pseudomonas syringae pv. syringae B728a.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=205918;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RG DOE Joint Genome Institute;  
RA Chain P., Larimer P., DiBartolo G., Copeland A., Lykidis A., Trong S.,  
RA Nolan M., Goldsman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,  
RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.;  
RT "Comparison of two complete genome sequences of Pseudomonas syringae  
pv. syringae B728a and pv. tomato DC3000."  
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).  
SQ SEQUENCE 2502 AA; 277032 MW; A0069019F516BF0A CRC64;  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RA Loper J.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=B728a;  
RA Feil H., Feil W.S., Lindow S.E.;  
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
DR EMBL; CP000075; AAY39235.1; -; Genomic DNA.  
SQ SEQUENCE 2502 AA; 277032 MW; A0069019F516BF0A CRC64;  
[3]

Query Match 30.3%; Score 3948; DB 2; Length 2502;  
Best Local Similarity 36.0%; Pred. No. 9.1e-201;  
Matches 962; Conservative 474; Mismatches 922; Indels 314; Gaps 68;  
Qy 8 LNKISPTRDGQMTLADLQVLSFSEL-----RKIFDDQLSWGEARHLYHETIEQKQN 60  
Db 3 LNRLLS--RIKRLSARAGLDSLAAQALISPAFRNKLSTSGLSPREVDLYESARDERAA 60  
Qy 61 RLLEARLFRANPOLSGAIRLGIE-RDSVSRSDYEMFGARSSSFVKPGSVASMSFAGYL 119  
Db 61 LIYEKCLLARSPLLNKAVRLGINPPDASLRDYEQFGRASSYTLPGSVSMFSPAAL 120  
Qy 120 TELYREAKDLHFSSAYHLDNRDPDLADLTLSQSNMDEITSLTSLNELLE----- 171  
Db 121 SALYRNARGLYPESPHIDKRRPDLKGLSLQSNMKNKVSALSLSNEVMTLAGESLKL 180  
Qy 172 HITRKTGSDSALMESLSTYRQAIDTPPHOPEYETIRQVIMTHDSTLSALSNNPEVMGOAE 231



Db 181 DVTAEGQG---SVLEWLTSTYLSGSTPYHHPHARLRQSOIQKDPKPKOLAANPRVTGLFS 237  
Qy 232 GASILAILANISPELYNLTBEITEKNADALFAQNFSENIITPENFASOWIAKYIGLELS 291  
Db 238 GATWAGMADPISPEYALITBEVYTAGSADLEYARNFG-TIDPETLITPQSLRRYGLSDE 296  
Qy 292 EVQKYLGMQLONGYSTSAYVDNISTGLVNNESKLEAYKIKTRVKTDYDKNINVFDMY 351  
Db 297 EVALFSTDYANG-BEGABEYINLLVTRIGD-----KIYRQSGHPTNTLNAVWLLP 347  
Qy 352 EGNQOFFTRANKVUSREFGATLRKNAGPSGIVGSL-----SGPLTANTNPK-----SNY 400  
Db 348 QADGTWLLKSF-----SDAQGYDIFQVRVDTADAIYEENQEWGDFVPGQSY 396  
Qy 401 LSNISDSSEYKNG---VKIYAYRYSTSTSATNQGGGIFTFESYPLTIFALKINKAIRLCIT 457  
Db 397 TVSIEASDLLQGEFTMSLRHYNGSYVGVSYNNVIIDSF--QFYVLKINKAIRLWRA 454  
Qy 458 SGLSPNELQTVRSDNAGIINDSVLTKVFTLTFYSHRYALSFDDAQVLNGSVINQYADD 517  
Db 455 TGLHPKSLVIVNSVNPVN-ITDETQLQLFQVQRCVQRYGVEPEALVLCGLLSNNSYD 513  
Qy 518 DSVSHFRLNTPPLKGGKIFEADGNTVSIIDP-DEEQSTFARSALMRGIVGNSGELYQLGK 576  
Db 514 DNQSLFDQVNSPPLNGETIAPSTTSINLLPTNATDNTVEKAILKRALNDDVGLFALLR 573  
Qy 577 LAGVLDAQNTITLSVFISSLYRLTLARVHQLTVNELCMLYGLSPFNGKTTASISGEL 636  
Db 574 IFDPAASSGVLTNLKNLSAMVALSRWALHALSVTELEQL-----LKAAGL 620  
Qy 637 PRLVI-----WLYQV---TOWLFEAITTEAIWLLCTPBFSGNISPEISNLLNLL 683  
Db 621 PRLTSELESTQLWSGLQKVDLSLTQWLSNRKLSVADLELLTRPTPTQAAATEISALLEEL 680  
Qy 684 RPSISEDMAQSHRELQAEIIPFAATLHLASPDMAHYIILLTNDNTPGGLDIAGMTL 743  
Db 681 KGVISAH-PDADTLEKRLTALLPVLVSSLPSPAPVAESVLAWANGLOPABEWTVDQFW-- 737  
Qy 744 VLKESLANET--TOLVOFCHVMAQLSLSVQTLRLSEALSVLVTSIGFAVLGAKNQPAQ 801  
Db 738 -----DGAETDDSDSVAFYGLAQLALIYHATGINPOAFSLFVATPARLLG----PAPE 787  
Qy 802 -----HNIDTFLPSYRPHOWINGNPGSDTDLMLRQOTLTPADRLASVWGLDIDSMVTQAM 856  
Db 788 VAVLPRSLATVQALCNFSAWLSKLDGASAVLAAPVADTLTPADILARALNDDATFEQAT 847  
Qy 857 VSAGV-----QLOCHQDINTVLQWIDVASALHTMPSVIRTLVNTRYVTALNKAESNL 909  
Db 848 ROAYANTQAASETKLSAWSEIDVVLQW-----AALSTTFGV--TPVNIQGLLALNYTAGN 901  
Qy 910 PSWDEWOTLAENMEAGLSQOQAOTLADYTABRLSSVLCNWFNLANTQPBGSLSHRDDIYS 969  
Db 902 PPWDDWRVADAFTAGLTPYETKMEALASGLSALCYLLKSGMTAQLANNSREGLYQ 961  
Qy 970 YFLIDNQVSSAIKTRTLAEIAGIQLYINRALNRIEPPNADVS--TRQFTDW-TVNNR 1026  
Db 962 YLLDLNLNGPQWTSRVAEIAVSQTFIORTLSAAESQGTVDKAAVTQGFPTDWERYNQR 1021  
Qy 1027 YSTWGGVSRVLYPENYIDPTQRIQOTRMDDELLENISQSLSRDVEDAFKTYLTRPET 1086  
Db 1022 YSTWAGAAKVVYPENYDPTVRLQSGMMNTMLQTLGQAQLNTDVTGDAFTNLSNSEE 1081  
Qy 1087 VADLKVSAYHDNVNSNTGLTFWGVOTRENLPYVWRNVDTS-RMOAGELAAANAKWTK 1145  
Db 1082 VANLRVSGYHDNLDVHEGKTYFFIGTNSQVREFFYRSADGRSEDQQLAANATDWRK 1141  
Qy 1146 IDTAVNPYKDAIRPVIFRERLHLIWEKEEVA---KNGT--DPVETVDRFTLKLAFLRHD 1200  
Db 1142 IECAAPWGDGICRPVYKSRSLYLCWLERKQDVTTPPKADGSPGD-KMFD-YAINISYLRD 1199  
Qy 1201 GSWAPASWYDITTOVE--AVTDKRPDTERLALASGFGQEDTLLVFVYKTKGYSDFGGS 1258  
Db 1200 GNWTSPIIIDVTEENGPAAGKPP-----GIYCTSFDAETTMVALLYRKADSVADQLPV 1254

Qy 1259 NKNVAGMTIYGDGSPFKKMENTALSRYSQLKNTFDIIHTQGNLDRKASRYRFAQDFEVPAS 1318  
Db 1255 QADMRLVYIYEDMASDQROSSAETFLSSMK--YELDPTEGVWVNNKYTH----- 1301  
Qy 1319 LMGSAIGDDSLTWENGINPQIITSKYSSDNLAITLHNAAFTRYDGSNVIR---NKQI 1375  
Db 1302 ---GLSLGESVHLGERNY-DISVSYGS-----YTLDVNGSEAVAKISIRKI 1346  
Qy 1376 SAM-----KLTVGDGK-----SOVGNAFIIANTVKHYGGYS 1407  
Db 1347 TLVPRIPNPLPLLYEYSSPAETFFELRALSDDGKIWFVROGGKIYLLALSILTW----- 1401  
Qy 1408 LGGPITVYNTKKNYIASVOGHLMNADYTRRLILTPVENNYIARLPEFFPSPNTILNTVPT 1467  
Db 1402 -GVPLSVM-----VYGDHARLLVR--EDAVSAVTF-----VLA 1433  
Qy 1468 VGSNKTSPDKCSYAVDGNNSQGFQIF-----SSYQ-----SSGWL-----DIDTGIN 1510  
Db 1434 VYNDPFDMLCP-----SPAGFTVMQLVTGSYQFDTPPQMMANPPYFSPFDL----- 1482  
Qy 1511 NTDIKITVMAGSKTHTFTASDHIASLPANSFDMPTFKPLEIDAS-----SLAPT 1561  
Db 1483 ---LQLANVFDVNGQAFLAGDYVSALPPFDISGMVFPFDVSNVISKEMGVLPSTDVAFS 1539  
Qy 1562 NNIAPLDIVFETKAKGRVLGKIKOTLSVKRVNYPEDILFLRTHSGAQYMLGVYRIR 1621  
Db 1540 PSIGVD-----SDDR---IQWIVPVRNASDSKNIKIKNTPKAYMEFACRTR 1588  
Qy 1622 LNTLLASQLVSRANTGDTILTMTQRLPEPPLGEGFPANFVLPKYDPAEHGDEMFH 1681  
Db 1589 LNTLFARQLVERAAGIDTILSYETQEIQEPQLGAGFTTLTPHKGVHGDEPMVKIY 1648  
Qy 1682 IGNVGNTGRQPYYSGLMSDTSMTLFPVYAE-GYVMEHGVRVLGVGYKITDY--NTW 1738  
Db 1649 YNYFLAVNDSYLANSGSLSATSTTVNLFIYPDGGWFTTNEVHLRLQYKEADYNGLSQ 1708  
Qy 1739 ESAPFFYDETKQQFVLINDADHDSGMTQOQGIKVIKIKYKGFNLVSIATGYSAQMDPNSAS 1798  
Db 1709 SVWLLYAPDTGVATV---SRPGTNSLSEIVK-----SVTVKSRSRTVPMDFSGAN 1755  
Qy 1799 ALYTWELFYTPMCMCFORLLOEKQPDENATQINVYVNPAGYIVNGEIAPIWNCHEPLEET 1858  
Db 1756 ALYTWELFYTPMMAAQRFLOEQFTLADPWLRYSPSGYVVRGQHVDRNWNVPLOED 1815  
Qy 1859 TSWANPLDADPDAVAQNDDPMHYKIATFFMLLQDLILRGDMAYRELTRDALNEAKMYV 1918  
Db 1816 TSWNDAPLKAVDPAVAQNDDPMHYKVATFMRALDILLIARGDAAYRKLERDYLAEAKVYS 1875  
Qy 1919 RTLELLGDEPEDYSGSQWAAAPSLSGAASQTVQAAAYQOBLTWL-----GRGGVSVKNLRTAN 1973  
Db 1876 QALNLLGQPYIRANALWAEPSLGEASS---EALAEQHLTVLSLLRDRGRVETLKSMASTK 1932  
Qy 1974 SLV--GLFLPEYNPALTDYQOTLRLPLNLRHNLSDGQPLSLAIYAEPTPKALLTSMV 2031  
Db 1933 TAAASALFLPEINVMQGYMLTLRQMYNLRHNLTDGQPLLLPLFAKPAKALLNAV 1992  
Qy 2032 QASQGSQAVLPGT--LSLYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDDELTLTLQ 2090  
Db 1993 ARESGSGDLEPVTSLPWRFEPLDSDAGLVFQLIQFNAVGLVLERQDAESLALLQ 2052  
Qy 2091 GMEIATOSIRIQORTVDEVDADIADVLAESRSQAQRLEKYQOQLYDEDEINHGEQRAMSLD 2150  
Db 2053 GTELMASTIQOEGTRELBAEKVLSKDSQARRPDSYRSLYDEDEINSRERLSSLE 2112  
Qy 2151 AAQOSLAGQVLSAEGVADLVPNVFGIACGSGRWGAALRASASVMSLSATASQVSADKI 2210  
Db 2113 SAKSLGAKAAYIAAALDLPANI FGLANGMKFPGVGTWAGLGISISAENLMDMSRI 2172  
Qy 2211 SRSEAYRRRREWEIQRDNADGEVKQMDAQLESKIRREAAQOMOVEYQETQQAHTQAOLE 2270  
Db 2173 SQEIIYRRRREWEIQRNABGEIQQIEAQLASLEVRRESTELQKHALEMQOQQAQOLD 2232

```
Qy 2271 LLOKFTNKALYSWMRGKLSAIYYQFFDLTQSFCLMAEALRRELTDNGVTFIRGGAWNG 2330
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2233 FLOTKFNSALYSWMRGKLSAIYYQFFDLTQSFCLMAEALRRELTDNGVTFIRGGAWNG 2291
Qy 2331 TTAGLMAGETILLNLAMEKVLWDERDRALERTVTSLSAQFYQA--LSSDNFNITEKLTQ 2388
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2292 TWAGLTCGEGMLNLAQLETARMKWSKRALBITVTSLSADPYRSTLAETDQFELSAAVSA 2351
Qy 2389 FLREGKNGVAGSNELKL-SNRQTEASVRLSDLKIFSDYPESLGNTRQLQKVSVTLPALV 2447
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2352 LL-NGDTPPEGSARVRLDESGALTASITLADLNIVDDYPGGLGQDRRIKQVSVSLPALL 2410
Qy 2448 GPVEDIRAVLNY-GGSIVMPRGCSAIALSHGVNDSGOFMLDNDSRVLPPEGISVNDSGS 2506
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2411 GPYQDIOAVLNYTGGNVLPGCCDNMAISRGVNDSGQFPDPNDPRMLPPEGADIRE-GS 2469
Qy 2507 LTLSPFQDTRQKALLESLSDIILHIRYTRS 2538
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2470 MIISFPQAEKQKALLESNDIILHINYTRS 2501

RESULT 15
Q693A5 YEREN
ID Q693A5 YEREN PRELIMINARY; PRT; 2168 AA.
AC Q693A5
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Putative insecticidal toxin complex protein.
GN Name=tcbA;
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T83;
RA Temant S.M., Skinner N.A., Joe A., Robins-Browne R.M.;
RT "Yersinia enterocolitica biotype 1A insecticidal toxin complex
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY647257; AAT90756.1; -; Genomic DNA.
SQ SEQUENCE 2168 AA; 245531 MW; 6767A56CCF1BE064 CRC64;

Query Match 22.7%; Score 2965.5; DB 2; Length 2168;
Best Local Similarity 31.0%; Pred. No. 1.8e-148;
Matches 814; Conservative 367; Mismatches 847; Indels 601; Gaps 67;

Qy 29 SFSLRKIFDQLSGEARHLYHETIEOKNNRLLERIFTRANPQLSGAIFRLGIERDSV 88
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
22 SLEEVYTAQLDTHLSLSEIQVLYNADKKRKAQAQQRSSNRGNPQLQOATLBITADLQ 81
Qy 89 SRSYDEMFGA-----RSSSFVKPGSVASMPSPAGYLITELYREAKDLHPSSAYHL 138
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
82 DTOSTDASGAMFYNSDMLERHIVANGDIDASRFSPPAAYITELYKNAKSLHKPTSLYI 141
Qy 139 DNRPPDLADLTLSQNMDETSTLTLSNELLEHITRTKGGSDALMESLSTYQAIDTP 198
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
142 DKRRPDIAELLSDQNQDSVSTLDSNKLITHTIKQKSGDKS--VTEILKTH----NIP 195
Qy 199 YHQPYETIROVIMTHDSTLSAL---SRNPEVMQAGCASILAILANISPELYNLTBEI 254
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
196 FNLEFEKQVQLKARKTTAAQWKKHINEPT-----TTLGLG--GLSYCYQLTWP 246
Qy 255 TEKNADALFAQNFSENITPENFASQSWIAKYGLSEVQKYLQMLONGYSDSTSAVDN 314
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
247 QLDNWPILPLE-----MARYVELPPSVIDSEVHLLPSGKDN-----283
Qy 315 ISTGLVNVNESKLEAYKILTRUKTDDYDKNINFDLMYEGNNOFFIRANPKVSRREGATLR 374
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
284 -----ARRTR-----HOLLM-----LH 296
Qy 375 KNAGPSGIVGSLGSLPIANTNFKSNLYLSIDSEYKNGVKIYAVRYTSTSATNQGGGIF 434
```

```
Db KTTG-----300
Qy 435 TFSFYPTITIPALKUNKAIRLCRLTSGLSGPNELQITVRSDNAQGIINDSVLTQVFTLFSH 494
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301 -FSLATLTLFVLK-----DGHFSTDCVLQHLHTNIYQQ 332
Qy 495 RYALSPDDAQVLYNGSVINQYADDDSVSHFNRLFNTPLPGKIPEADGNTVSIIDPEEQST 554
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
333 RYGLISATDA-----IT 343
Qy 555 FARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRLTLARVHQLTVNEL 614
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
344 LVRQDKAG-----VPSTPVAALLABCHQLTLAEL 373
Qy 615 CMLYGLSPFNGKTTASLSSGSELPRLVITLWLVQVOTWLTAEAITTEAILWLLCTPESGNISP 674
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
374 SALRPL-----IADANSTDSYKFIETVNTQMLNHHQSVADLVNMTSTTDTQTQTA 426
Qy 675 EISNLNANLRPSISEDMAQSHNRELOAEIILAPFTAATLHLASPDWARYILLWTNLRPGG 734
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
427 EIEHLVTTLOANVTR--SELDRLVQLRFELSSP-LAQLLEL-TPAATTFLLYWMESG-RPNG 481
Qy 735 LDIAGFMTLV--LKESLNANETTOLVQFCHVMAQLSLSVQTLRLSEALSVLVTSGFVL 792
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
482 VKIEDIWQKIINLDTSKPFKVPASVIAFCHQLAQLAQITKKLHLTDEALLQLARTDLGLT 541
Qy 793 GAKNQAPAGQHNIDTLFSLYRPHQWINGLGNPGS--DTLDMLRQOQTLTADRLASVMGLDIS 850
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
542 IVTN--NKFTADGLMELSYRDLMALSEQDSNADTI-LVEQKPL--QELLEMFSEVLL 595
Qy 851 MVTQAMVSAGVNLQ----CWODINTVLQWIDVASALHTMPSVIRTLVNIR--VVTALNK 904
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
596 ETPSGDAKGFNNLRYPDFDFKNIESQYWOQIAKTLHLPVDSILPLSDMHQDLQATRL 655
Qy 905 AESNLPSDEWQTLAENMEAGLSTQQAQTLADYTAERLSSVLCNWFANLQIPEGVSLHSR 964
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
656 IDS-----LAETAFGTLPASDRSTSIDILAEKQSEALSRYYL-TIPR---LKTR 701
Qy 965 DDLYSYFLIDNOVSSAKITRLAEIAGIOLYINLRNL--IEPNARADVSTROFFTDWT 1022
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702 DDYISYLLIDNQVTFPLIKTIGAIBAIASLOLYVNRQNGNEGVOKSLKRGFFTDWD 761
Qy 1023 V-NNRYSTWGGVSRVLYVPENYIDPTORIGOTRMMDELLENISQSKLSRDTVEDAFKTYL 1081
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
762 LYNKRYSSWSGINSQITFPENYIDPALRIGQTSMDTILQSIQSLSADTLEDAFKTYL 821
Qy 1082 TRPFTVADLKVYSAYHDVNSNTGLTWFGQTRNLPEYVYWRNVDISRMQAGELAAANAWK 1141
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
822 TEPEEIANLSVITGYHDHESIDQGLTYFGBSLTSPPTYVWRTVNHOTFDEGKFPATAWS 881
Qy 1142 EWTIKIDTAVNPYKDAIRPVIPIRELRHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHG 1201
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 EWNKITTAAMPYNQLIRPFRNLGRLLYLWLREERENNRKPPDGAKMICYELKLSHLRYDG 941
Qy 1202 SWSAPWSVDITTOVEAVTDKPPDTERLALAAASQFGEDTLVVFVYKTKSYSDSGGSN-- 1259
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
942 TWSVPFNFV-----EISDTYFQEGCSLHCTASSPTELHVCLYQKADSYSEKSIKLM 996
Qy 1260 --KNVAGMTIYGDGSFKKQEN--TALSRYSQLKNTFDIIHTQGNLDLVKAS--YRPAQDF 1313
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
997 IIDEMMGLT-----KQENKEKLKSWVSSLSNEFDTL-----TERRISNFYSGGKKV 1042
Qy 1314 EYPASLNGMSAIGDSSLTWMENGIPOIT-SKYS-SDNLATITLHNAAPTVDYDGSNVIR 1371
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1043 SVRNTINEKPKI-----ISTTHIGNPKFSAHSQYGFQVDVNIINSTNKDITFOFKPFPVEIN 1098
Qy 1372 NKQISAMKLTGVDGKSVQGNAFIIANTVKHYGGYS-----DLGGP-----ITV- 1414
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1099 NOVIDLGDISRLADPSLLESALKKHGENKQYLAYSREINFDLTGDNALTCNISFIMVTIP 1158
Qy 1415 YNKTKNYIASVOGHLMN---ADYTRRLILITPV-----ENNYIARL-FE 1453
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 1159 YEEYDNY--EDRNRVLYNSGRDVKVYLSAPVPIKMSFNDEERFNKLEKNRKYSELSFA 1216  
Qy 1454 PPFSENTILNTVFTVGS-----NKTSDFKKCSYAVDGN-----NSQGFQ 1492  
Db 1217 GDYGNRHSULTPSIGSKPNELIFAYVNDLLDEDLIEAYIDDLNFMGSMWDRNSDSYQ 1276  
Qy 1493 IF-----SSVQSSGWLDDIDGINTNDIKITVMAGSKTHPTASD-HIASLPANSP 1541  
Db 1277 YYPEKNITITPKTPASIESLITTSVGNKVV-----SGPYPTANNKKSADILVHKW 1330  
Qy 1542 DAMPYTPKPLEIDASSLAFTNNAIDLVFETKAKDGRVLGKI KOTLSVKRVNYPEDIL 1601  
Db 1331 DOPIELTLPVASEST-----DIEIHSYHSRNEVSGSMIOTIPIITVDPVSQNYL 1382  
Qy 1602 FLRETHSGAQYMLG---VYIRILNTLLASQVSRANTGIDITILMTQRLPEPPLGEGF 1658  
Db 1383 ALKTNEAQAYLEWDKQKHQVRLNTLFAKELVREADSGIDNLSPTQLIFEPLDGEN- 1441  
Qy 1659 FANFVLPKYDPAEHGDERWFKIHIGNVGNGTRQRPYSGMLS DTSSETMTLFPVYAEYY 1718  
Db 1442 ----- 1441  
Qy 1719 MHEGVLGVGYKITDYDNTWESAPFYFDETKQFVLIINDADHDSGMTQGGIVKNIKYYKG 1778  
Db 1442 -----NNHAANQS----- 1449  
Qy 1779 FLNVSITAGYSAPMDFNSASALYWEFYFYPMMCQRLQEQKQFDEATOWINYVNPAG 1838  
Db 1450 -----VDFSGANGLYFWELFYFYPMLVMQRLQEQSPDLATHMLSYVFN- 1494  
Qy 1839 YVNGEIAPIWNCPLBETTSWNPANPLDADPDVAQNDPMHYKIATFMRLLDOLILRG 1898  
Db 1495 ---TNPSTP--WNTRPLKEDTAWNAPLDSTNPDAVAQADPMHYKLS TFMRLLDOLILRG 1549  
Qy 1899 DMAYRELTRDALNEAKMYVTRLELLGDEP-EDYGSQQWAAAPSLSGAASQTVOAAQQDL 1957  
Db 1550 DYAYHQOQORDTLAEAKMMYVQANNLLGGEPTRTDSRW----- 1588  
Qy 1958 TMLGRGGVSKNLRTANSLVG-LFLPEYNPALTDYQTLRLRLNLRHNLSDIGOPLSIAI 2016  
Db 1589 -----VHLPLKEAVELADHLFLPQONTKLQDYQILKLRVFNLRNNLSIDGQPLMLPQ 1641  
Qy 2017 YABPTDPKALITSMVOASQGSVALPG--TSLYRFPVWLERTRNLVAQLTQPTSLISM 2074  
Db 1642 FTAPADAKALMNTAVASSLGGIA-LPSHIKLSLQRPFPVLESAMVQGLIQFGSSLSGV 1700  
Qy 2075 AEHDDADELTTLLOQGMELATQSTIRIQORTVDEVDADI AVLAEBSRRAQNRLEKYQOLY 2134  
Db 1701 IERDAEKMAVLMQQTGTELLAQSVLMQOKNLEELHETHTALSAALAGAEGRYKHNDLY 1760  
Qy 2135 DEDINHGQRAMSLIDAAAGOSLACQVLSIAEGVADLVPNVFGACGSGRWGAALRASAS 2194  
Db 1761 QENINAGENSAIALRATAGFITGTTQALYMTGAALDMVPNIYGVMSVGARYGAIANAIAI 1820  
Qy 2195 VMSLSATASQYSADKISRSEAYRRRRQBEWEORDNADGEVQMDAQLESKIRREAAQMO 2254  
Db 1821 GSSIASTASTLAADGISTTEMYRRRRQBEWEISRNNAAEIKQITAAQOETLKIRSTSAQL 1880  
Qy 2255 VEYQETQQAHTQAQLELLQKFTNKALYSWMRGKLSALYYOFPDITQSFCLMAQEARRE 2314  
Db 1881 VSYLETQQAQHTQALMNKFTNQALYSWMRSRLSALYFYDITVSRCLQAQYGYQWE 1940  
Qy 2315 LTDNGVTPIRGAMNGTTAGLMAGETLLINLAEMEKVWLERDERALEVTRTVSLAQFYQA 2374  
Db 1941 TLDI--TSFIKPGAMQGTYAGLLCGEALMLNLATLEAAVQSMQARELEIERTLSLAELYQ 1999  
Qy 2375 LSSDNFNUTEKLTQPLRGK--GNVYASGNEKLKSNRQIBASVRLSDKIFSDYPE--SL 2430  
Db 2000 IPTSGFDLATAISAALSASNKVSQTCGNGNVISIKDDILSLSFISALNLAQDYPKIMGL 2059  
Qy 2431 GNTRQLKQVSVTLPALVGPEDIRAVANYGGS-IWMPRGCSAIALSHGVNDSGGOFMLDFN 2489  
Db 2060 GDKRPFQIQISVSLPALLGYSQVAVLSYTGTDATFAKGCDAIALSGMNDSGLFDLDFN 2119

Qy 2490 DSRYPFEGISVNDSGSLTSLSPDATDRQKALLESLSDIILHIRYTIRS 2538  
Db 2120 DSNYLPFEGIDIEDSGFVLRFPTNSDKQKALQSLSDIILHIRYTIRS 2168

Search completed: February 16, 2006, 21:42:47  
Job time : 366.476 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 251.656 Seconds  
(without alignments)  
4431.220 Million cell updates/sec

Title: US-10-754-115-34  
Perfect score: 13043  
Sequence: 1 MYSTAVLNKISPRDGTQTM.....KALLESLSIIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21:.\*  
1: geneseqp1980s:.\*  
2: geneseqp1990s:.\*  
3: geneseqp2000s:.\*  
4: geneseqp2001s:.\*  
5: geneseqp2002s:.\*  
6: geneseqp2003as:.\*  
7: geneseqp2003bs:.\*  
8: geneseqp2004s:.\*  
9: geneseqp2005s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13043	100.0	2538	8	Adn61385 Xenorhabd
2	13043	100.0	2538	8	Adr20371 Recombina
3	13043	100.0	2538	8	Adr21514 Xenorhabd
4	13043	100.0	2538	8	Adr21500 Xenorhabd
5	5496.5	42.1	2523	8	Adr20365 Recombina
6	5496.5	42.1	2523	8	Adr21494 Xenorhabd
7	5148.5	39.5	2522	2	Aay33729 Photorhab
8	5043	38.7	2516	2	Aaw56572 Toxin Tcd
9	5043	38.7	2516	4	Aab72609 Photorhab
10	5043	38.7	2516	5	Abg32651 P. lumine
11	5043	38.7	2516	8	Adp18614 Photorhab
12	5043	38.7	2516	8	Adr21575 Photorhab
13	5043	38.7	2516	9	Adz72157 P. lumine
14	5043	38.7	2517	4	Aab72611 Modified
15	5043	38.7	2537	4	Aab72614 TcdA toxi
16	5003	38.4	2516	2	Aaw17899 Photorhab
17	4986.5	38.2	2504	2	Aaw56557 Toxin Tcd
18	4986.5	38.2	2504	4	Aab72610 Photorhab
19	4986.5	38.2	2504	5	Abg32654 P. lumine
20	4986.5	38.2	2504	4	Adr21539 Photorhab
21	4986.5	38.2	2505	4	Aab72612 Modified
22	4976	38.2	2534	9	Aeb47821 Native Xp
23	4966.5	38.1	2504	2	Aaw17871 Photorhab
24	4914.5	37.7	2526	6	Abm70229 Photorhab

25	4802	36.8	2381	8	ADP18620	Adp18620 Photorhab
26	4802	36.8	2381	8	ADR21543	Adr21543 Photorhab
27	4763.5	36.5	2466	6	ABM70225	Abm70225 Photorhab
28	4583.5	35.1	2376	4	AAY97694	Aay97694 SepA prot
29	4574	35.1	2499	8	ADP18616	Adp18616 Photorhab
30	4572	35.1	2499	8	ADR21542	Adr21542 Photorhab
31	4520.5	34.7	2499	6	ABM70221	Abm70221 Photorhab
32	4491	34.4	2222	6	ABM70220	Abm70220 Photorhab
33	4433	34.0	2177	8	ADP18618	Adp18618 Photorhab
34	3303	25.3	1849	2	AAW56573	Aaw56573 Toxin Tcd
35	3278.5	25.1	1844	2	AAW18302	Aaw18302 Photorhab
36	3278.5	25.1	1844	2	AAW56558	Aaw56558 Toxin Tcd
37	3264	25.0	1849	2	AAW17900	Aaw17900 Photorhab
38	2855	20.4	1273	9	ABE47816	Aeb47816 Partial n
39	1643	12.6	1205	8	ADN61341	Adn61341 Paenibaci
40	1632	12.5	573	2	AAW56559	Aaw56559 Toxin Tcd
41	1622	12.4	573	2	AAW18303	Aaw18303 Photorhab
42	1615	12.4	579	2	AAW56574	Aaw56574 Toxin Tcd
43	1611	12.4	579	2	AAW18301	Aaw18301 Photorhab
44	1605.5	12.3	1370	6	ABM67445	Abm67445 Photorhab
45	1595.5	12.2	1565	6	ABM69065	Abm69065 Photorhab

## ALIGNMENTS

### RESULT 1

ADN61385  
ID ADN61385 standard; protein; 2538 AA.

XX ADN61385;

XX 01-JUL-2004 (first entry)

XX Xenorhabdus nematophila Xwi XptA2 toxin complex protein.

XX Cry; toxic; lepidopteran pest; toxin complex; insecticide; xwi; xptA2.

XX Xenorhabdus nematophila.

XX WO2004002223-A2.

XX 08-JAN-2004.

XX 27-JUN-2003; 2003WO-US020082.

XX 28-JUN-2002; 2002US-0392633P.

XX 21-JAN-2003; 2003US-0441647P.

XX (DOWC ) DOW AGROSCIENCES LLC.

XX Bintrim SB, Bevan SA, Zhu B, Merlo DJ;

XX WPI; 2004-082821/08.

XX Screening a culture of Paenibacillus isolate for Cry protein or toxin complex protein, useful for controlling lepidoptera, comprises obtaining DNA or protein from the culture and assaying the presence of the gene or protein.

XX Example 10; SEQ ID NO 49; 220pp; English.

XX The invention relates to a novel method for screening a culture of a Paenibacillus isolate for a gene encoding a protein selected from a Cry protein that is toxic to a lepidopteran pest and a toxin complex protein. The method comprises obtaining DNA from the culture and assaying the DNA for the presence of the gene or obtaining a protein produced by the culture and assaying the presence of a protein that indicates the presence of the gene in the isolate. The method of the invention has insecticide applications and may be useful for screening Paenibacillus sp. for toxin complex (TC)-like genes and proteins which may themselves be used to enhance or potentiate the activity of a stand-alone Xenorhabdus toxin protein. The method may also be useful for screening

CC Paenibacillus sp. and others for insecticidal thiaminase genes and  
CC proteins for controlling insects, particularly lepidopterans. The current  
CC sequence is that of the xenorhabdus nematophila Xwi XtpA2 toxin complex  
CC protein of the invention.

Q	SQ	Sequence	2538	AA;
		Query Match	100.0%;	Score 13043; DB 8; Length 2538;
		Best Local Similarity	100.0%;	Pred. No. 0;
		Matches 2538; Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MYSTAVLLNKISPTRDGQTMTLADLQYLSFSSELKIFDDQLSGWGARHLYHETIIBQKKN	60	
Db	1	MYSTAVLLNKISPTRDGQTMTLADLQYLSFSSELKIFDDQLSGWGARHLYHETIIBQKKN	60	
Qy	61	RLLEARIPTRANPOLSGAIRLGIERSDSVRSYDENFMFGARSSSFVKPFGSVASMFSPAGYLT	120	
Db	61	RLLEARIPTRANPOLSGAIRLGIERSDSVRSYDENFMFGARSSSFVKPFGSVASMFSPAGYLT	120	
Qy	121	ELYREAKDLHFSSSAYHLNDRRPLADLTLSQSNMDTEISTLTLSNELLEHETIRKTCGD	180	
Db	121	ELYREAKDLHFSSSAYHLNDRRPLADLTLSQSNMDTEISTLTLSNELLEHETIRKTCGD	180	
Qy	181	SDALMESLSTYRQAITDPYHOPYETIROVIMTHDSTLSALSRNPEVMGQAEASLLAILA	240	
Db	181	SDALMESLSTYRQAITDPYHOPYETIROVIMTHDSTLSALSRNPEVMGQAEASLLAILA	240	
Qy	241	NISPELYNILTEETEKNA DALFAONFSENITPENFASQSWITAKYGYELSEVQKYLGM	300	
Db	241	NISPELYNILTEETEKNA DALFAONFSENITPENFASQSWITAKYGYELSEVQKYLGM	300	
Qy	301	QNGYSDSSTSAVDNISTGLVNNESKLRAYKTRVKTDDYDKNINFDLMYEGNNQFPFR	360	
Db	301	QNGYSDSSTSAVDNISTGLVNNESKLRAYKTRVKTDDYDKNINFDLMYEGNNQFPFR	360	
Qy	361	ANFKVSRREFGATLRKNAGPSGIVGSLSGPLIANTNFKSNYLSNISDSSEYKNGVKIYAYRY	420	
Db	361	ANFKVSRREFGATLRKNAGPSGIVGSLSGPLIANTNFKSNYLSNISDSSEYKNGVKIYAYRY	420	
Qy	421	TSSSTSATNQGCGGITPFESYPLTIFALKNKAIKRLCLTSGLSNPELQTVRSNDAQIIND	480	
Db	421	TSSSTSATNQGCGGITPFESYPLTIFALKNKAIKRLCLTSGLSNPELQTVRSNDAQIIND	480	
Qy	481	SVLTAKVFTYTLFYSHRYALSFDQAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGKIFEAD	540	
Db	481	SVLTAKVFTYTLFYSHRYALSFDQAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGKIFEAD	540	
Qy	541	GNTVSIIDPDEQSQTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFIISLRYL	600	
Db	541	GNTVSIIDPDEQSQTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFIISLRYL	600	
Qy	601	TLLARVHQLTVNELCMLYGLSPFNCKTTASLSGGELPRLVIWLQYQVQWLTAEAITTEAI	660	
Db	601	TLLARVHQLTVNELCMLYGLSPFNCKTTASLSGGELPRLVIWLQYQVQWLTAEAITTEAI	660	
Qy	661	WLLCTPFPSPGNISPEISNLLNLRPSISEDMAQSHNRELQABILAPPTAATHLASPDMA	720	
Db	661	WLLCTPFPSPGNISPEISNLLNLRPSISEDMAQSHNRELQABILAPPTAATHLASPDMA	720	
Qy	721	RYIILLWTNLRPGGLDIAGFMTLVLKESINANETTLQVQFCHVMAQLSUSVQTLRLSEAE	780	
Db	721	RYIILLWTNLRPGGLDIAGFMTLVLKESINANETTLQVQFCHVMAQLSUSVQTLRLSEAE	780	
Qy	781	LSVLVIVISGFAVLGAKNQAGQHNIDTLFSLYRFHQWINGLGNPGSDTLDMLRQOQTLTADR	840	
Db	781	LSVLVIVISGFAVLGAKNQAGQHNIDTLFSLYRFHQWINGLGNPGSDTLDMLRQOQTLTADR	840	
Qy	841	LASVWGLDISHVYQAMVSAGVNOLOCQDINTVLQWIDVASAHTMPSVIRTLVNIYVT	900	
Db	841	LASVWGLDISHVYQAMVSAGVNOLOCQDINTVLQWIDVASAHTMPSVIRTLVNIYVT	900	
Qy	901	ALNKAESNLPDWDEWQTLAENNEAGLSTQQAQTLTADYTAERLSVSLCNWFLANIQPEGVS	960	

Qy	2041	LPGTL	SLYRFPV	MLER	TRNI	VAQ	TQ	FGT	SLLS	MAEHDD	DADEL	TTL	LLQ	QGMEL	ATQ	SIR	2100																																										
Db	2041	LPGTL	SLYRFPV	MLER	TRNI	VAQ	TQ	FGT	SLLS	MAEHDD	DADEL	TTL	LLQ	QGMEL	ATQ	SIR	2100																																										
Qy	2101	IQORT	VEVD	AD	TA	VLA	ESRR	SQ	NR	LEK	YQ	Q	YD	ED	IN	HGEOR	AMSL	LLDAA	AGOS	LAGQ	2160																																						
Db	2101	IQORT	VEVD	AD	TA	VLA	ESRR	SQ	NR	LEK	YQ	Q	YD	ED	IN	HGEOR	AMSL	LLDAA	AGOS	LAGQ	2160																																						
Qy	2161	VLST	AE	GV	AD	LV	PN	V	FL	AC	GS	R	WG	AA	L	R	AS	V	MS	L	S	AT	AS	Q	Y	S	AD	K	I	S	R	S	E	A	Y	R	R	R	2220																				
Db	2161	VLST	AE	GV	AD	LV	PN	V	FL	AC	GS	R	WG	AA	L	R	AS	V	MS	L	S	AT	AS	Q	Y	S	AD	K	I	S	R	S	E	A	Y	R	R	2220																					
Qy	2221	QEW	E	I	Q	R	N	D	N	A	D	G	E	V	K	O	M	D	A	O	E	S	L	K	I	R	R	E	A	O	Q	Y	E	T	Q	O	A	H	T	O	A	O	E	L	L	O	R	K	T	N	K	A	2280						
Db	2221	QEW	E	I	Q	R	N	D	N	A	D	G	E	V	K	O	M	D	A	O	E	S	L	K	I	R	R	E	A	O	Q	Y	E	T	Q	O	A	H	T	O	A	O	E	L	L	O	R	K	T	N	K	A	2280						
Qy	2281	LYS	WM	R	G	K	L	S	A	I	Y	Q	F	D	L	T	Q	S	P	C	L	M	A	Q	E	A	L	R	R	E	L	T	D	N	G	V	T	F	R	G	G	A	N	G	T	T	A	G	L	M	A	G	E	T	2340				
Db	2281	LYS	WM	R	G	K	L	S	A	I	Y	Q	F	D	L	T	Q	S	P	C	L	M	A	Q	E	A	L	R	R	E	L	T	D	N	G	V	T	F	R	G	G	A	N	G	T	T	A	G	L	M	A	G	E	T	2340				
Qy	2341	LLN	L	M	E	M	E	K	V	M	L	E	R	D	E	R	A	L	E	V	T	R	T	V	S	L	A	O	F	Y	Q	A	L	S	S	D	N	F	N	L	T	E	K	L	T	Q	F	U	R	E	G	K	N	V	G	A	S	2400	
Db	2341	LLN	L	M	E	M	E	K	V	M	L	E	R	D	E	R	A	L	E	V	T	R	T	V	S	L	A	O	F	Y	Q	A	L	S	S	D	N	F	N	L	T	E	K	L	T	Q	F	U	R	E	G	K	N	V	G	A	S	2400	
Qy	2401	GNE	L	K	L	S	N	R	O	I	E	A	S	V	R	L	S	D	L	K	I	F	S	D	Y	P	E	S	I	G	N	T	R	O	L	K	O	N	S	V	T	L	P	A	L	V	G	P	E	D	I	R	A	V	L	N	Y	G	2460
Db	2401	GNE	L	K	L	S	N	R	O	I	E	A	S	V	R	L	S	D	L	K	I	F	S	D	Y	P	E	S	I	G	N	T	R	O	L	K	O	N	S	V	T	L	P	A	L	V	G	P	E	D	I	R	A	V	L	N	Y	G	2460
Qy	2461	GSI	V	M	P	R	G	C	S	A	I	A	L	S	H	G	V	N	D	S	G	O	F	M	L	D	F	N	D	S	R	V	L	P	P	E	G	I	S	V	N	D	S	G	S	I	T	L	S	F	P	D	A	T	R	O	K	A	2520
Db	2461	GSI	V	M	P	R	G	C	S	A	I	A	L	S	H	G	V	N	D	S	G	O	F	M	L	D	F	N	D	S	R	V	L	P	P	E	G	I	S	V	N	D	S	G	S	I	T	L	S	F	P	D	A	T	R	O	K		

RESULT 2  
ADR20371  
ID ADR20371 standard; protein; 2538 AA.

04-NOV-2004 (first entry)

Recombinant pDAB2097 cosmid open reading frame 7 protein SeqID 20.

toxin complex; TC: insect; pest control; exochitinase; insecticide;

pesticide; agriculture.

**Xenorhabdus nematophila.**

WO2004067750-A2.

12-AUG-2004.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

07-JAN-2005; 2005MO-US00000313.

ZI-JAN-2003; 2003US-044111/F.

(DOWC) DOW AGROSCIENCE

BILLI III SB, Mico  
Schafer BW, Beva

WPI; 2004-581006/56.

Claim 1; SEQ ID NO 20; 133pp; English.

This invention relates to novel nucleic acid molecules isolated from *Xenorhabdus* strain Xwi that encode toxin complex (TC) proteins. Specifically, it refers to proteins that are toxic to insects and other pests, such that they can be used to control pests in an agricultural environment to reduce crop loss and improve crop quality. The present invention further describes an excitinase enzyme obtainable from the Xwi strain that can be used to intoxicate or poison infected insect populations. As such, these toxins function as orally active insect control agents that can disrupt or deter insect growth and/or feeding, which may or may not cause death of the insect but will work to avoid insect-induced plant damage. The toxin is delivered via transgenic plant expression, formulated protein compositions, sprayable protein compositions or a bait matrix. Accordingly, these compositions exhibit insecticidal and pesticide activities. This polypeptide sequence is a *Xenorhabdus nematophila* open reading frame protein sequence encoded by DNA from the recombinant pDAB2097 cosmid insert given in an exemplification of the invention.

Sequence 2538 AA;

### Every Match

```
QY 721 RYILLWTDNLPPGLDIAAGFMTLVLSKESINANETTQLVQFCHVAQLSLSVQTLRLSEAE 780
Db 721 RYILLWTDNLPPGLDIAAGFMTLVLSKESINANETTQLVQFCHVAQLSLSVQTLRLSEAE 780
QY 781 LSVLVISGFAVLGAKNQAPAGHNDITLFSLFYRHHQWINGLNGPAGSDTLDMRLQOQTLTADR 840
Db 781 LSVLVISGFAVLGAKNQAPAGHNDITLFSLFYRHHQWINGLNGPAGSDTLDMRLQOQTLTADR 840
QY 841 LASVGLDLSMVTQAMVSAGVNLQCCWQDINTVLQWIDVASALHTMPSVIRTLVNIYVT 900
Db 841 LASVGLDLSMVTQAMVSAGVNLQCCWQDINTVLQWIDVASALHTMPSVIRTLVNIYVT 900
QY 901 ALNKAESNLPSWDEWQTLAENWEAGLSLQOQTLADYTAERLSSVLCNWFPLANTQPEGV 960
Db 901 ALNKAESNLPSWDEWQTLAENWEAGLSLQOQTLADYTAERLSSVLCNWFPLANTQPEGV 960
QY 961 LHSRDDLYSYFLIDNOVSSAIKTTRLABAIAIGIOLYINRALNRIEPNARADVSTROQFTD 1020
Db 961 LHSRDDLYSYFLIDNOVSSAIKTTRLABAIAIGIOLYINRALNRIEPNARADVSTROQFTD 1020
QY 1021 WTVNNRSTWGGVSRSLVYPENYIDPTQRIIGOTRMDELLENISQSKLSRDTVEDAFKTY 1080
Db 1021 WTVNNRSTWGGVSRSLVYPENYIDPTQRIIGOTRMDELLENISQSKLSRDTVEDAFKTY 1080
QY 1081 LTRPETVADIAKVSAYHDNVNSNTGLTWFGQGTRENLPYYWRNVDISRMQAGELAAW 1140
Db 1081 LTRPETVADIAKVSAYHDNVNSNTGLTWFGQGTRENLPYYWRNVDISRMQAGELAAW 1140
QY 1141 KEWKIDTAVNPYKDAIRPVIFRERLHLIWIWEKEVAKNGTDPVETYDRFTLKLAFLRHD 1200
Db 1141 KEWKIDTAVNPYKDAIRPVIFRERLHLIWIWEKEVAKNGTDPVETYDRFTLKLAFLRHD 1200
QY 1201 GSWAPMSYDITTOVEAVTDKPKDTERLALAASGQGEDTLLVVFYKTKSYSPFGGSNK 1260
Db 1201 GSWAPMSYDITTOVEAVTDKPKDTERLALAASGQGEDTLLVVFYKTKSYSPFGGSNK 1260
QY 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN 1320
Db 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN 1320
QY 1321 MGSAGDDSLTWMNGNIPOITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380
Db 1321 MGSAGDDSLTWMNGNIPOITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380
QY 1381 TGVDGSKSYGNAFIANTVKHYGYSYDLGGPITVYNTKNYIASVQGHLMNADYTRRLIL 1440
Db 1381 TGVDGSKSYGNAFIANTVKHYGYSYDLGGPITVYNTKNYIASVQGHLMNADYTRRLIL 1440
QY 1441 TPVENNYIARLPEPPFSNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFIFSSYQSS 1500
Db 1441 TPVENNYIARLPEPPFSNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFIFSSYQSS 1500
QY 1501 GWLDDITGINNTDIKITWAGSKTHTFTASDHIAASLPANSFPDAMPYTFKPLEIDASSLAF 1560
Db 1501 GWLDDITGINNTDIKITWAGSKTHTFTASDHIAASLPANSFPDAMPYTFKPLEIDASSLAF 1560
QY 1561 TNNIAPLDIVFETKAKDGRVLGKIQTLSVKRVNYPNPEDILFLRETHSGAQYMQLVGYRI 1620
Db 1561 TNNIAPLDIVFETKAKDGRVLGKIQTLSVKRVNYPNPEDILFLRETHSGAQYMQLVGYRI 1620
QY 1621 RLNTLLASQLSVRANTGIDTILTMETORLPEPPLGEGFFANFVLPKYDPAEHWDERWEFKI 1680
Db 1621 RLNTLLASQLSVRANTGIDTILTMETORLPEPPLGEGFFANFVLPKYDPAEHWDERWEFKI 1680
QY 1681 HIGNVGNTGQPYYSGLMSTSETSMTLFVPYAEGYMHEGVLGVGYQKITYDNTWES 1740
Db 1681 HIGNVGNTGQPYYSGLMSTSETSMTLFVPYAEGYMHEGVLGVGYQKITYDNTWES 1740
QY 1741 AFFYFETKQQFVLINDADHDSGMTQOQGIKXNKKYKGFNLVSIATGYSAPMDFNASAL 1800
Db 1741 AFFYFETKQQFVLINDADHDSGMTQOQGIKXNKKYKGFNLVSIATGYSAPMDFNASAL 1800
```

```
QY 1801 YYWELFYTPMWCFORLLQEKQFDEATQWVINYVNPAGYIVNGEIAPIWNCRPLEETTS 1860
Db 1801 YYWELFYTPMWCFORLLQEKQFDEATQWVINYVNPAGYIVNGEIAPIWNCRPLEETTS 1860
QY 1861 WNANPLDAIDPDVAQNDDPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNANPLDAIDPDVAQNDDPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
QY 1921 LELLDGDEPDYGSQOAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSILVGLFL 1980
Db 1921 LELLDGDEPDYGSQOAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSILVGLFL 1980
QY 1981 PEYNPALTDYQWTLRLRLFLNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGSASV 2040
Db 1981 PEYNPALTDYQWTLRLRLFLNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGSASV 2040
QY 2041 LPGLTSLYRPPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTTLLIQQGMELATQSI 2100
Db 2041 LPGLTSLYRPPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTTLLIQQGMELATQSI 2100
QY 2101 IQORTVDEVDADIIVLAESRRSAQNRLKYLQYDEIDINHGEORAMSLDDAAAGOSLAGO 2160
Db 2101 IQORTVDEVDADIIVLAESRRSAQNRLKYLQYDEIDINHGEORAMSLDDAAAGOSLAGO 2160
QY 2161 VLSIABGVADILVFNVPFGLACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRERR 2220
Db 2161 VLSIABGVADILVFNVPFGLACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRERR 2220
QY 2221 QEWEIFQRDNADGEVKQMDAQLLESKIRREAAQMVQYQETQQAHTQAQLELLQKRFITKA 2280
Db 2221 QEWEIFQRDNADGEVKQMDAQLLESKIRREAAQMVQYQETQQAHTQAQLELLQKRFITKA 2280
QY 2281 LYSWMRGKLSAIYQFDELITQSFCLMAQEALRRELITONGVTFFIRGGAWNGTTAGLMAGET 2340
Db 2281 LYSWMRGKLSAIYQFDELITQSFCLMAQEALRRELITONGVTFFIRGGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMEKVWLERDERALEVTRVSLAQFYQALSSDNFNLTEKLTQFLREGKGNVGAS 2400
Db 2341 LLLNLAEMEKVWLERDERALEVTRVSLAQFYQALSSDNFNLTEKLTQFLREGKGNVGAS 2400
QY 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
Db 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
QY 2461 GSIWMPRGCSAIALSHGVNDGSGQFMDPNDSDRYLPFEGISVNDGSLTSLSPDATDROKA 2520
Db 2461 GSIWMPRGCSAIALSHGVNDGSGQFMDPNDSDRYLPFEGISVNDGSLTSLSPDATDROKA 2520
QY 2521 LLESLSDIILHIRTIRS 2538
Db 2521 LLESLSDIILHIRTIRS 2538
RESULT 3
ADR21514
ID ADR21514 standard; protein; 2538 AA.
XX ADR21514;
AC ADR21514;
XX 04-NOV-2004 (first entry)
XX Xenorhabdus strain Xwi cosmid pDAB2097 ORF7 deduced protein SEQ ID NO:34.
XX toxin; insect; insecticidal; transgenic; pest control; cosmid.
XX Xenorhabdus nematophila.
XX WO2004067727-A2.
XX 12-AUG-2004.
XX 07-JAN-2004; 2004WO-US000394.
XX XX
```



PR 21-JAN-2003; 2003US-0441723P.  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX Hey JD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
 XX Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
 XX WPI; 2004-580999/56.  
 XX DR N-PSDB; ADR21499.  
 XX Controlling or inhibiting an insect, useful for pest control, comprises  
 PT contacting the insect with effective amounts of a Protein A, a Protein B,  
 PT and a Protein C.  
 XX Claim 1; SEQ ID NO 34; 368pp; English.  
 XX The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from xptA1wi,  
 CC xptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdC, TcdC1wi, XptB1xb,  
 CC PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wi, XptC1xb, PptC1  
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC the deduced protein of ORF7 (xptA2wi) of cosmid insert pBAB2097, obtained  
 CC from Xenorhabdus nematophila strain Xwi.  
 XX Sequence 2538 AA;  
 Query Match 100.0%; Score 13043; DB 8; Length 2538;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MYSTAVLLNKISPTRDGQMTLADLQYLSFSELKIFDDQLSWGSEARHLYHETIBQKXN 60  
 DB 1 MYSTAVLLNKISPTRDGQMTLADLQYLSFSELKIFDDQLSWGSEARHLYHETIBQKXN 60  
 QY 61 RLLEARIETFRANPOLSGAIRLGIERSVSRSDYDEMFGARSSSFVKPGSVASMFSPAGYLT 120  
 DB 61 RLLEARIETFRANPOLSGAIRLGIERSVSRSDYDEMFGARSSSFVKPGSVASMFSPAGYLT 120  
 QY 121 ELYREAKDLHFSSAYHLDRRPDLADLTLSQSNMDEITSLTSLNELLLEHITRKTGDD 180  
 DB 121 ELYREAKDLHFSSAYHLDRRPDLADLTLSQSNMDEITSLTSLNELLLEHITRKTGDD 180  
 QY 181 SDALMESLSTYRQADITPHQYETIROVIMTHDSTLSALSRLNPEVMQAGASLAILA 240  
 DB 181 SDALMESLSTYRQADITPHQYETIROVIMTHDSTLSALSRLNPEVMQAGASLAILA 240  
 QY 241 NISPELYNLTBEITEKADALFAQNFSENITPENFASQSWITAKYGLSEVQKYLGM 300  
 DB 241 NISPELYNLTBEITEKADALFAQNFSENITPENFASQSWITAKYGLSEVQKYLGM 300  
 QY 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDKNINFDLMYEGNNOFFIR 360  
 DB 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDKNINFDLMYEGNNOFFIR 360  
 QY 361 ANFKYSRREGATLRKNAGPSGIVGSLGPLIANTNFKSNLYLSNIDSSEYKNGVKIYAYRY 420  
 DB 361 ANFKYSRREGATLRKNAGPSGIVGSLGPLIANTNFKSNLYLSNIDSSEYKNGVKIYAYRY 420  
 QY 421 TSSTSATNQGGIIFTFESYPLTIFALKLNKAIKRLCLTSGLSNPELQTVIRSDNAQGIIND 480

DB 421 TSSTSATNQGGIIFTFESYPLTIFALKLNKAIKRLCLTSGLSNPELQTVIRSDNAQGIIND 480  
 QY 481 SVLTAKVFTYFYSHRYALSPDDAQVYLVNGSVINQVADDDSVSHFNRLFNTPPLKGIKIFRAD 540  
 DB 481 SVLTAKVFTYFYSHRYALSPDDAQVYLVNGSVINQVADDDSVSHFNRLFNTPPLKGIKIFRAD 540  
 QY 541 GNTVISIDPDEEQSTFARSALMRGIGVNSGELYQJGKLAGVLDQAQNTITLSVPVTSLSYRL 600  
 DB 541 GNTVISIDPDEEQSTFARSALMRGIGVNSGELYQJGKLAGVLDQAQNTITLSVPVTSLSYRL 600  
 QY 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASLSSGELPRVLVWLYQVTLQWLTAEAITTEAI 660  
 DB 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASLSSGELPRVLVWLYQVTLQWLTAEAITTEAI 660  
 QY 661 WLLCTPESGNIPEISNLNLRPSISEDMAQSHNRELOAEIILAPFIAATLHLASPDMA 720  
 DB 661 WLLCTPESGNIPEISNLNLRPSISEDMAQSHNRELOAEIILAPFIAATLHLASPDMA 720  
 QY 721 RYILLWTDLNLRPGGLDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780  
 DB 721 RYILLWTDLNLRPGGLDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780  
 QY 781 LSVLVISGFVILGAKNQAPAGQHNDITLPSLYRFHQWINGLGNPGSDTILDMRLQOQLTADR 840  
 DB 781 LSVLVISGFVILGAKNQAPAGQHNDITLPSLYRFHQWINGLGNPGSDTILDMRLQOQLTADR 840  
 QY 841 LASVGLDLSMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900  
 DB 841 LASVGLDLSMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900  
 QY 901 ALNKAESNLPSEDEWQTLAENWEAGLSLTOOQTLADYTAERLSSVLCNWFNLANIPEGVS 960  
 DB 901 ALNKAESNLPSEDEWQTLAENWEAGLSLTOOQTLADYTAERLSSVLCNWFNLANIPEGVS 960  
 QY 961 LHSRDDLYSYFLIDNQVSSAIKTTLAEALAGIOLYINRNLNRIEPNARADVSTQFFTD 1020  
 DB 961 LHSRDDLYSYFLIDNQVSSAIKTTLAEALAGIOLYINRNLNRIEPNARADVSTQFFTD 1020  
 QY 1021 WTVNNRSTWGGVSRVLYPENYIDPTQRIQOTRMDELLENISQSKLSRDTVEAFKTY 1080  
 DB 1021 WTVNNRSTWGGVSRVLYPENYIDPTQRIQOTRMDELLENISQSKLSRDTVEAFKTY 1080  
 QY 1081 LTRFETVADLVKVSAYHDNVNSNTGLTFVFGQTRNLPEYWRNVDSIRMOAGELAAANAW 1140  
 DB 1081 LTRFETVADLVKVSAYHDNVNSNTGLTFVFGQTRNLPEYWRNVDSIRMOAGELAAANAW 1140  
 QY 1141 KEWKIDTAVNPYKDAIRPVIKRELRHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200  
 DB 1141 KEWKIDTAVNPYKDAIRPVIKRELRHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200  
 QY 1201 GSWAPMSYDITTOVEAVTDKKPOTERLALAASGQGEDTLLVFFYKTKGSYDFGGSNK 1260  
 DB 1201 GSWAPMSYDITTOVEAVTDKKPOTERLALAASGQGEDTLLVFFYKTKGSYDFGGSNK 1260  
 QY 1261 NVAGMTIYGDGSPKKMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN 1320  
 DB 1261 NVAGMTIYGDGSPKKMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN 1320  
 QY 1321 MGSAGDSDSLTWENGNIPQITTSKYSSDNLAITLHNAAFVTRYDGSNGVNRNKOISAMKL 1380  
 DB 1321 MGSAGDSDSLTWENGNIPQITTSKYSSDNLAITLHNAAFVTRYDGSNGVNRNKOISAMKL 1380  
 QY 1381 TGVDGSKSYGNAFIANTVKHYGYSYDLGGPITVYNTKNTKNTYASVQGHLMNADYTRRLIL 1440  
 DB 1381 TGVDGSKSYGNAFIANTVKHYGYSYDLGGPITVYNTKNTKNTYASVQGHLMNADYTRRLIL 1440  
 QY 1441 TPVENNYARLFEPPFSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSGQFOIFSSYQSS 1500  
 DB 1441 TPVENNYARLFEPPFSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSGQFOIFSSYQSS 1500  
 QY 1501 GMLDITGINTNDIKITVWAGSKTHTFTASDHIAASLPANSFDMPTYTFKPLEIDASSLAF 1560  
 DB 1501 GMLDITGINTNDIKITVWAGSKTHTFTASDHIAASLPANSFDMPTYTFKPLEIDASSLAF 1560

QY 1561 TNNIAPLDIVPETKAKDGRVLGKIQTLSVKRVNYPEDILFLRETHSGAQMVLGVYRI 1620  
Db 1561 TNNIAPLDIVPETKAKDGRVLGKIQTLSVKRVNYPEDILFLRETHSGAQMVLGVYRI 1620  
QY 1621 RLNTLLASQLVSRANTGIDTILTMETQRLPEPPLGEGFFANFVLPKYDPAEHDGDERWFKI 1680  
Db 1621 RLNTLLASQLVSRANTGIDTILTMETQRLPEPPLGEGFFANFVLPKYDPAEHDGDERWFKI 1680  
QY 1681 HIGNVGGNTGRQPYYSGLMSTSETSMTLFVPYABGYTHGVRGLGVGYKITDYNTWES 1740  
Db 1681 HIGNVGGNTGRQPYYSGLMSTSETSMTLFVPYABGYTHGVRGLGVGYKITDYNTWES 1740  
QY 1741 AFFYFDETKQOQVLLINDADHDSGMTQOGIVKNIKKYKGLVNSIATGYSAPMDFNSASAL 1800  
Db 1741 AFFYFDETKQOQVLLINDADHDSGMTQOGIVKNIKKYKGLVNSIATGYSAPMDFNSASAL 1800  
QY 1801 YWELFYFTYPMWCFORLLQEQFDEATQWINVYVNPAGYIYNGEIAPIWNCRPLEETTS 1860  
Db 1801 YWELFYFTYPMWCFORLLQEQFDEATQWINVYVNPAGYIYNGEIAPIWNCRPLEETTS 1860  
QY 1861 WNAFLDAIDPDVAQNPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920  
Db 1861 WNAFLDAIDPDVAQNPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920  
QY 1921 LELIGDEPEDYGSQOQWAAPSLSGARASQTVQAAQQDLTMLGGGVSKNLTANSIYGLFL 1980  
Db 1921 LELIGDEPEDYGSQOQWAAPSLSGARASQTVQAAQQDLTMLGGGVSKNLTANSIYGLFL 1980  
QY 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIQPLSLAIYAEPTDPKALLTSMVQASQGSVAV 2040  
Db 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIQPLSLAIYAEPTDPKALLTSMVQASQGSVAV 2040  
QY 2041 LPGTSLSYRFPVMLERTNLVAQLTQFGTSLLSMAEHDADDELITLLQQGMELATQSIIR 2100  
Db 2041 LPGTSLSYRFPVMLERTNLVAQLTQFGTSLLSMAEHDADDELITLLQQGMELATQSIIR 2100  
QY 2101 IQQRTVDEVDADIAVLASRRSAONLEKYQOQYDEIDNHGEORAMSLDDAAAGOSLAGO 2160  
Db 2101 IQQRTVDEVDADIAVLASRRSAONLEKYQOQYDEIDNHGEORAMSLDDAAAGOSLAGO 2160  
QY 2161 VLSIAEGVADLVPNVFLGACGSRGWAALRASASVMSLSATASQYSADKISRSEAYRRRR 2220  
Db 2161 VLSIAEGVADLVPNVFLGACGSRGWAALRASASVMSLSATASQYSADKISRSEAYRRRR 2220  
QY 2221 QEWETQRNADGEVKQMDAQLESKIRREAAQMVQEYQETOQAHTQAOLELLQRKFTNKA 2280  
Db 2221 QEWETQRNADGEVKQMDAQLESKIRREAAQMVQEYQETOQAHTQAOLELLQRKFTNKA 2280  
QY 2281 LYSWMRGKLSAIYYQFPLTQSFCMLAQEALRRELTDNGVTFIRGGANWGTAGIMAGET 2340  
Db 2281 LYSWMRGKLSAIYYQFPLTQSFCMLAQEALRRELTDNGVTFIRGGANWGTAGIMAGET 2340  
QY 2341 LLLNLAEMKVLDERALEVTRTVSLAQFYQALSSDNFNLTCLKTQFLREGKGNVGAS 2400  
Db 2341 LLLNLAEMKVLDERALEVTRTVSLAQFYQALSSDNFNLTCLKTQFLREGKGNVGAS 2400  
QY 2401 GNEKLSNRQIEASVRLSDLKIFSDYPESLGNTRQLKQSVTLPALVGPYEDIRAVLNYG 2460  
Db 2401 GNEKLSNRQIEASVRLSDLKIFSDYPESLGNTRQLKQSVTLPALVGPYEDIRAVLNYG 2460  
QY 2461 GSIWVPRGCSAIALSHGVNDGQFMLDFNDSRYLPEGISVNDGSLTSLSPDATDRQKA 2520  
Db 2461 GSIWVPRGCSAIALSHGVNDGQFMLDFNDSRYLPEGISVNDGSLTSLSPDATDRQKA 2520  
QY 2521 LLESLSDIILHIRTYS 2538  
Db 2521 LLESLSDIILHIRTYS 2538

RESULT 4  
ADR21500  
ID ADR21500 standard; protein; 2538 AA.

XX ADR21500;  
XX 04-NOV-2004 (first entry)  
XX Xenorhabdus strain Xwi cosmid pDAB2097 ORF7 deduced protein SEQ ID NO:20.  
XX toxin; insect; insecticidal; transgenic; pest control; cosmid.  
XX Xenorhabdus nematophila.  
XX WO2004067727-A2.  
XX 12-AUG-2004.  
XX 07-JAN-2004; 2004WO-US000394.  
XX 21-JAN-2003; 2003US-0441723P.  
XX (DOWC ) DOW AGROSCIENCES LLC.  
XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
XX Ni W, Zhu B, Merlo DU, Apel-Birkhold PC;  
XX WPI; 2004-580999/56.  
XX N-PSDB; ADR21499.  
XX Controlling or inhibiting an insect, useful for pest control, comprises  
XX contacting the insect with effective amounts of a Protein A, a Protein B,  
XX and a Protein C.  
XX Claim 12; SEQ ID NO 20; 368pp; English.  
XX The invention relates to a novel method for controlling or inhibiting an  
XX insect comprising contacting the insect with effective amounts of a  
XX Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
XX is encoded by a naturally occurring gene or has an amino acid sequence  
XX that differs from the product encoded by a naturally occurring gene only  
XX by truncation or by conservative amino acid changes. Protein A is a 230-  
XX 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
XX species, has stand alone insecticidal activity, and has an amino acid  
XX sequence at least 40% identical to a sequence selected from XprAlwi,  
XX XpCA2wi, TcdA, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin  
XX complex potentiator having an amino acid sequence at least 40% identical  
XX to a sequence selected from TcdB1, TcdB2, TcdC, XpC1wi, XpC1xb,  
XX PpC1(orfs), or SepC. Protein C is a 90-120 kDa toxin complex potentiator  
XX having an amino acid sequence at least 35% identical to a sequence  
XX selected from TccC1, TccC2, TccC3, TccC4, TccC5, XpC1wi, XpC1xb, PpC1  
XX (orf 6 long), PpC1 (orf 6 short), and SepC. Also claimed is a transgenic  
XX plant or plant cell that produces a Protein A, a Protein B, and a Protein  
XX C. The method is useful for pest control. The present sequence represents  
XX the deduced protein of ORF7 (XpCA2wi) of cosmid insert pDAB2097, obtained  
XX from Xenorhabdus nematophila strain Xwi.  
XX SQ Sequence 2538 AA;  
Query Match 100.0%; Score 13043; DB 8; Length 2538;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MYSTAVLLNKISPRDGTMTLADLQVLSFSELAKIFDDQLSWGAEARHLHYETIEQKNN 60  
Db 1 MYSTAVLLNKISPRDGTMTLADLQVLSFSELAKIFDDQLSWGAEARHLHYETIEQKNN 60  
QY 61 RLLEARIIFRANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120  
Db 61 RLLEARIIFRANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120  
QY 121 ELYREAKDLHFSSAYHLNDRRPDLADLTLSQSNMDTEISTLTLSNELLLEHITRKTCGD 180  
Db 121 ELYREAKDLHFSSAYHLNDRRPDLADLTLSQSNMDTEISTLTLSNELLLEHITRKTCGD 180  
QY 181 SDALMESLSTYRQAIDTPYHPYETIRQVIMTHDSTLSALSRNPVEMQAGCASILLAILA 240

Db 181 SDALMESLTYRQADTDPHQYETIRQVIMTHDSTLSALSBNPEVMQAGASLAILA 240  
Qy 241 NISPELYNLTBEITEKADALFAQNFSENIIPENFASQSWIAKYVGLSESVQKYLGL 300  
Db 241 NISPELYNLTBEITEKADALFAQNFSENIIPENFASQSWIAKYVGLSESVQKYLGL 300  
Qy 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDXNINYPDLMEGNNQFFIR 360  
Db 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDXNINYPDLMEGNNQFFIR 360  
Qy 361 ANFKYSRFBGATLRKNAGPSGIVGSLGPIANTNPKSNLYLSNISDSSEYKNGVKIYAYRY 420  
Db 361 ANFKYSRFBGATLRKNAGPSGIVGSLGPIANTNPKSNLYLSNISDSSEYKNGVKIYAYRY 420  
Qy 421 TSSTSATNQGCGIFTFESYPLTFIPALKLNKALRLCLTSGLSNPLQTVIRSDNAQGIIND 480  
Db 421 TSSTSATNQGCGIFTFESYPLTFIPALKLNKALRLCLTSGLSNPLQTVIRSDNAQGIIND 480  
Qy 481 SVLTQVFTYTLFYSHYALSFDDAQVLNGSVINQYADDSVSHFNRLFPNTPPLKGIPEAD 540  
Db 481 SVLTQVFTYTLFYSHYALSFDDAQVLNGSVINQYADDSVSHFNRLFPNTPPLKGIPEAD 540  
Qy 541 GNTVSDIDPEEOSTFARSALMRLGVLNGSELVOLKLAGVLDQANTITLSVFVSSLYRL 600  
Db 541 GNTVSDIDPEEOSTFARSALMRLGVLNGSELVOLKLAGVLDQANTITLSVFVSSLYRL 600  
Qy 601 TLLARVHQLTVNELCWLGLSPFNKTTASLSSGELPRLVILWYQVOWLTAEAITTEAI 660  
Db 601 TLLARVHQLTVNELCWLGLSPFNKTTASLSSGELPRLVILWYQVOWLTAEAITTEAI 660  
Qy 661 WLLCTPEFSGNISPEISNLNLRPSISBDMQASHNRELQAEILAPFIAATLHLASPDMA 720  
Db 661 WLLCTPEFSGNISPEISNLNLRPSISBDMQASHNRELQAEILAPFIAATLHLASPDMA 720  
Qy 721 RYLLWTONLRPGGDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLSEAE 780  
Db 721 RYLLWTONLRPGGDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLSEAE 780  
Qy 781 LSVLVISGFVAVLGAKNQAGQHNIDTLFSLRPHQWINGLNGPSGSDTLDMRLQOFTLTADR 840  
Db 781 LSVLVISGFVAVLGAKNQAGQHNIDTLFSLRPHQWINGLNGPSGSDTLDMRLQOFTLTADR 840  
Qy 841 LASVGLDISMVTQAMVSAGVNLQCVQDINTVLQWIDVASALHTMPSVIRTLVNIYVT 900  
Db 841 LASVGLDISMVTQAMVSAGVNLQCVQDINTVLQWIDVASALHTMPSVIRTLVNIYVT 900  
Qy 901 ALNKAESNLPWDWEQTLAENWEAGLSTQOAGTLADYTAERLSSVLCNWFPLANTQPEGVS 960  
Db 901 ALNKAESNLPWDWEQTLAENWEAGLSTQOAGTLADYTAERLSSVLCNWFPLANTQPEGVS 960  
Qy 961 LHSRDDLYSYFLIDNQVSAIKTTRLABAIAQIQLYINRNLNRIEPPNARADVSTROFFTD 1020  
Db 961 LHSRDDLYSYFLIDNQVSAIKTTRLABAIAQIQLYINRNLNRIEPPNARADVSTROFFTD 1020  
Qy 1021 WTVNNRYSTWGVSVRLVYPENYIDPTQRIQGTMMDBELLENISQSKLSDRTVEDAFKTY 1080  
Db 1021 WTVNNRYSTWGVSVRLVYPENYIDPTQRIQGTMMDBELLENISQSKLSDRTVEDAFKTY 1080  
Qy 1081 LTRPETVADLVKVSAYHDNVNSNTGLTWVFGQGTRENLPYYWRNVDISRMQAGELAAANW 1140  
Db 1081 LTRPETVADLVKVSAYHDNVNSNTGLTWVFGQGTRENLPYYWRNVDISRMQAGELAAANW 1140  
Qy 1141 KEWKIDTAVNPKDAIRPVIIFRERLHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200  
Db 1141 KEWKIDTAVNPKDAIRPVIIFRERLHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200  
Qy 1201 GWSAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFYKTKGYSDFGGSNK 1260  
Db 1201 GWSAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFYKTKGYSDFGGSNK 1260  
Qy 1261 NVAGMTIYDGSFKOKENTALSRYSQLKNTFDI IHTQGNLDLVRKASYRPAQDFEVPASLN 1320

Db 1261 NVAGMTIYDGSFKOKENTALSRYSQLKNTFDI IHTQGNLDLVRKASYRPAQDFEVPASLN 1320  
Qy 1321 MGSAGDDSLTWENGNIPQITSKYSSDNLAITLHNAFTVRYDGSNGVIRNKQISAMKL 1380  
Db 1321 MGSAGDDSLTWENGNIPQITSKYSSDNLAITLHNAFTVRYDGSNGVIRNKQISAMKL 1380  
Qy 1381 TGVDGKSQYGNAPIIANTVKGHYGSDLGGPITVYNKTKNYIASVOGHLNADYTRRLJL 1440  
Db 1381 TGVDGKSQYGNAPIIANTVKGHYGSDLGGPITVYNKTKNYIASVOGHLNADYTRRLJL 1440  
Qy 1441 TPVENNYARLPEPFPSPNTILNTVFTVGSNKTSDFPKKCSYAVDGNNSOGFOIFSSYQSS 1500  
Db 1441 TPVENNYARLPEPFPSPNTILNTVFTVGSNKTSDFPKKCSYAVDGNNSOGFOIFSSYQSS 1500  
Qy 1501 GMLDITDGINNTDIKITWAGSKTHTFTASDHIAISLPANSFDPAMPYTPKPLEIDASSLAF 1560  
Db 1501 GMLDITDGINNTDIKITWAGSKTHTFTASDHIAISLPANSFDPAMPYTPKPLEIDASSLAF 1560  
Qy 1561 TNNIAPLDIVPETKAKOGRVLGKIQTLSVKRVNTPEDILFLRETHSGAQVQLGVYRI 1620  
Db 1561 TNNIAPLDIVPETKAKOGRVLGKIQTLSVKRVNTPEDILFLRETHSGAQVQLGVYRI 1620  
Qy 1621 RLNTLLASQLSVRANTGIDTILTMETQRLPEPPLGEGFPANFVLKYPDPAEHGDERWFKI 1680  
Db 1621 RLNTLLASQLSVRANTGIDTILTMETQRLPEPPLGEGFPANFVLKYPDPAEHGDERWFKI 1680  
Qy 1681 HIGNVGNTGRQPYYSGLMDSMTSETMTLFPVYAEYTMHGVGLGVGQKITYDNTWBS 1740  
Db 1681 HIGNVGNTGRQPYYSGLMDSMTSETMTLFPVYAEYTMHGVGLGVGQKITYDNTWBS 1740  
Qy 1741 AFFPDETQOQFVLINDADHDSGMTQOQIVNKKYKGLNVSIAATGYSAPMDFNSASAL 1800  
Db 1741 AFFPDETQOQFVLINDADHDSGMTQOQIVNKKYKGLNVSIAATGYSAPMDFNSASAL 1800  
Qy 1801 YWELFYTPMWCORLLOEQFDEATOWINVYNPAGYVINGETAPWNCRLPEETTS 1860  
Db 1801 YWELFYTPMWCORLLOEQFDEATOWINVYNPAGYVINGETAPWNCRLPEETTS 1860  
Qy 1861 WNANPLDAIDPDVAQNDDPMHYKIATFNRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920  
Db 1861 WNANPLDAIDPDVAQNDDPMHYKIATFNRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920  
Qy 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSVLGLFL 1980  
Db 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSVLGLFL 1980  
Qy 1981 PEYNPALTDYWQTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPPKALLTSMVQASOGGSV 2040  
Db 1981 PEYNPALTDYWQTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPPKALLTSMVQASOGGSV 2040  
Qy 2041 LPGTLSLYRFPVMLERTNLVAQLTQFCTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100  
Db 2041 LPGTLSLYRFPVMLERTNLVAQLTQFCTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100  
Qy 2101 IQORTVDEVDADIIVLAEBSRSANRLSKYQOOLYDEEDINHGEOQRAWSLLDAAAGSLAQ 2160  
Db 2101 IQORTVDEVDADIIVLAEBSRSANRLSKYQOOLYDEEDINHGEOQRAWSLLDAAAGSLAQ 2160  
Qy 2161 VLSIAEGVADIVPNVFGIACGSRWGAALRASASVMSISATASOYSAADKISRSEAYRRRR 2220  
Db 2161 VLSIAEGVADIVPNVFGIACGSRWGAALRASASVMSISATASOYSAADKISRSEAYRRRR 2220  
Qy 2221 QEWIQRDNADGKQMDAQLESKIRREAAQMOVYQETQQAHTQOALQLLQKFTNKA 2280  
Db 2221 QEWIQRDNADGKQMDAQLESKIRREAAQMOVYQETQQAHTQOALQLLQKFTNKA 2280  
Qy 2281 LYSWMRGKLSAIYYQFDFLTQSFCMAQEARLRELTNGVTFFIRGGAWNGTTAGLMAGET 2340  
Db 2281 LYSWMRGKLSAIYYQFDFLTQSFCMAQEARLRELTNGVTFFIRGGAWNGTTAGLMAGET 2340  
Qy 2341 LLLNLAEMEKVWLRDRERALSRTVSLAQFQALSSDNFNLTEKLTQFLREGKGNVGAS 2400  
Db 2341 LLLNLAEMEKVWLRDRERALSRTVSLAQFQALSSDNFNLTEKLTQFLREGKGNVGAS 2400



1072 TVEDAFKTYLTRFETVADLKVVSAHYNDVNSNTGLTWFGVGTRENLPEYIYWRNVDIRMQ 1131  
 1061 TVEDSFKNLTAFEDVANLQVIGYHDSINVNEGLTYLIGYSQTEPRYIYWRNVHOKCO 1120  
 1132 AGELAAWKEWTKIDTAVNPKDAIRPVIPERLHLIWEKEEVAKNCTDPVEYIDRPT 1191  
 1121 HQFAANAWGEWKIEIPINWQENIRPVYIKSRLLYLLWLEQKELNSESDEKIDITDIY 1180  
 1192 LKLAFLRHDSGSAFWSYDITTVQEAVIDKQDPTERLALAAAGFOCEDTLLVYVYKTGKS 1251  
 1181 LKLSHRYDGSNSPFPNVTDKIENLNKAS---IGMYCSDYEKDVIIIVPHEKDN 1237  
 1252 YSDFGSKNKGAMTYIGDGSFKKMENTALSYSOLKNTFDIHTQGNLDLVRKASVRFQA 1311  
 1238 YS--FNSLPAREGMITNPMTLISLTENDLD--AIVKSTLSBLDTRTE---YKVNQFAT 1290  
 1312 DP-----EVPASLNGSAG--DDSLTVMNGNI-----PQITSKYSSDN----- 1349  
 1291 DYLAEBYKESITTKNKLASFTGNI FDLSYISPGNGHINLTFFNPSMEINFSKGINYDEVKY 1350  
 1350 -LAITLHNAFTVRYDGSNGVNRKQISAMKLTGVGDGKSOYGNAPFIANTVHKYGYSDL 1408  
 1351 LLSWVEDETVILFDVRHDEM-----GKEE-----EVHYG----- 1382  
 1409 GGPITVYNTKNYIASVOGHLNADYTRRL-----ILTPVENNYIYARLPEF 1454  
 1383 -----TLDFIISID--LKNAEYFRVLMHLRTKEIPKSEIGVINGVINDYESNDAEF 1431  
 1455 PPSPTIILN-----TVFTVGNKTSDFK-----KCSYAVDGNNS 1488  
 1432 KLDITVILVDKONTGVWHTICESFTNDVSIIINMGNAIALFUREDPVCVYLCSIAATD---- 1487  
 1489 QGFQFSSVQSSGLDIDGIIINNTDKITVMAGSKTHT--FTASDHIASLPANSPAMPYT 1547  
 1488 --IKTASSMIEQ-----IDKNISFLKNGSDILVELNADHVASKPSHESHPMYD 1537  
 1548 FKPEIDASSLAFTNNAIPLDIVFETKAKDGRVLG-KIKQTLVSRKRVNYPEDILFLRET 1606  
 1538 FNQKVVDIEGYDI-----PLVSEFIKQPDGGYNDIVIESPIHKLKSKDTSNVLSHKM 1592  
 1607 HSGAQYMLGVRIIRLNTLLASOLVSRANTGIDTILTMETQRLPEPLGEGFPANFLPK 1666  
 1593 PSGTYMQIGPYTRTLNTLFSRKLAEANIGIDNVLSMETQNLPEPLGEGFYATFKLPP 1652  
 1667 YDPAEHDGERWPKIHNGVGGNTGQPYYSGLMDSSTSMTLFVYABGYMYHMGVRLG 1726  
 1653 YNKEEHDGERWPKIHNGIDGNSAKQPYEGMLSDI-ETITVTLFVYAKGYIIRGVRIG 1711  
 1727 VGYQKITDNTWESAFYPFDETQKQFVLINDADHDSGMTQOGIVKNIKKYKGLNVSAT 1786  
 1712 VGYKKIYDKSWESAFYFDETQKQFIFINDADHDSGMTQOGIVKNIKKYKGFHVVWK 1771  
 1787 GYSAPMDFNSASALYIYFELPYTPMCFORLLOEQFDEATOWINYVNPAGYIYNGETA 1846  
 1772 NNTPEMDFNAGANAIYFELPYTPMVFQRLLEQONFTETSTRLRYIWNYPAGYSVQGENQ 1831  
 1847 PWINCRPLEETTSNANPLDAIDPDVAQNDPMYKIAFMRLDOLLRCDMAIYREL 1906  
 1832 DTYWNRVPLEEDTSNANPLDSVDPDAQAQNDPMYKIAFMRLDOLLRCDMAIYREL 1891  
 1907 RDALNEAKWYVYRTLELLGDEPEDYGSQQAAPSLSGAASQTVQAAQQDLTMLRGGVYS 1966  
 1892 RDTLNEAKWYVYQALTLGDEPFYSLDNDWSPRLEEAASQTRHHYQHKMLQLQRAAL 1951  
 1967 KNLRTANSVLGLFELBEYNPALTDYQTLRLRLFNLRHNLSDIGQPLSLAIYAEPTDPKAL 2026  
 1952 PKRTANSITLFLPQINKLQGYQTLTQRLYLRHNLTDIGQPLSLSLYATPADPSML 2011  
 2027 LTSWQASQGSGLVPGTSLSLVRFVPMLERTNLVAQLTQFGTSLLSMAEHDADDELITL 2086  
 2012 LSAAITASQGGDLPHAVPMYRFPVILENAKWSQLIQFGNTLLSITERQDAALAEI 2071

2087 LLOQWELATQSIRIQORTVDEVDADIIVLAESRRSAQNRLKEKYQOOLYDEDINHGEORAM 2146  
 2072 LQTOGSELALOSIKMDQKVMABIDADKLALQSRHGAOSRFDSPNTLYDEDYNAGEKQAM 2131  
 2147 SLDDAAAGSLAGQVLSIAEGVADIVPNVFLGACGSRGWAALRASASVMSLSATASOYS 2206  
 2132 DLYLSSVLSVSTGTLHMAAADLVPNYIGFVAGSGRFGALFNASATGIEISASATRA 2191  
 2207 ADKISRSBAYRRRRQWEIQRDNADGEVKQMDAQLESKIRREAAQMVVEYQETOQAHTQ 2266  
 2192 ADKISOSEIYRRRRQWEIQRNABAEIKQIDAQLATLAVRREAAVLQKNYLETQOAOQTQ 2251  
 2267 AQLELLQKFTKALYSWMRGKLSAIYYQFPDLTQSFCLMAQEALRRELTDNGVTFIRG 2326  
 2252 AQLAFLOKSFNAALYNLGRGLSAIYYQFYDVLASCLMAEQTYOYELNNAAAHFKPG 2311  
 2327 ANWGTTAGLMAGETLLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSSDNFNLTXL 2386  
 2312 AWHGYTAGLAGEITMLNLAQMEKSYLEKDERALEVTRTVSLAEVYAGLTENSFILKDKV 2371  
 2387 TOFLREGKNGVASCNELKLSNRQIEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLPAL 2446  
 2372 TELVNAGSGSAGTTLNLNGVVEGTQOASLKLSDLIATDYPDGLGNTRRIKQISVTLPAL 2431  
 2447 VQPYEDIRAVLNGYSIVMPGCCSAIALSHGVNDSGQFMDFNDSRYLBPPEGISVNDSGS 2506  
 2432 LGPYQDVARILSYGSGTWMRCKAIAISHGMNDSGQFMDFNDAKYLPEGLPVADTGT 2491  
 2507 LTLSPDADTDRKALLESLSLIIHLIRYTIERS 2538  
 2492 LTLSPFGISGKKSULLSLIIHLIRYTIERS 2523

RESULT 6  
 ADR21494  
 ID ADR21494 standard; protein; 2523 AA.  
 XX  
 AC ADR21494;  
 XX  
 XX  
 DT 04-NOV-2004 (first entry)  
 XX  
 DE Xenorhabdus strain Xwi cosmid pDAB2097 ORF4 deduced protein SEQ ID NO:14.  
 XX  
 XX toxin; insect; insecticidal; transgenic; pest control; cosmid.  
 XX  
 OS Xenorhabdus nematophila.  
 XX  
 PN WO2004067727-A2.  
 XX  
 PD 12-AUG-2004.  
 XX  
 PF 07-JAN-2004; 2004WO-US000394.  
 XX  
 PR 21-JAN-2003; 2003US-0441723P.  
 XX  
 XX (DOWC ) DOW AGROSCIENCES LLC.  
 XX  
 PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
 XX  
 DR WPI; 2004-580999/56.  
 DR N-PSDB; ADR21493.  
 XX  
 PT Controlling or inhibiting an insect, useful for pest control, comprises  
 PT contacting the insect with effective amounts of a Protein A, a Protein B,  
 and a Protein C.  
 XX  
 PS Claim 1; SEQ ID NO 14; 368pp; English.  
 XX  
 CC The invention relates to a novel method for controlling or inhibiting an  
 CC insect comprising contacting the insect with effective amounts of a  
 CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence

that differs from the product encoded by a naturally occurring gene only by truncation or by conservative amino acid changes. Protein A is a 230-290 kDa toxin complex insect toxin that is derived from a first taxonomic species, has stand alone insecticidal activity, and has an amino acid sequence at least 40% identical to a sequence selected from XptAlwi, XptA2wi, TcdA, TcdB2, TcdM4, and TcdB. Protein B is a 130-180 kDa toxin complex potentiator having an amino acid sequence at least 40% identical to a sequence selected from TcdB1, TcdB2, TcaC, XptC1wi, XptC1xb, PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator having an amino acid sequence at least 35% identical to a sequence selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wi, XptC1xb, PptC1 (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic plant or plant cell that produces a Protein A, a Protein B, and a Protein C. The method is useful for pest control. The present sequence represents the deduced protein of ORF4 (XptAlwi) of cosmid insert pDAE2097, obtained from *Xenorhabdus nematophila* strain Xwi.

Sequence 2523 AA:

Query Match 42.1%; Score 5496.5; DB 8; Length 2523;

Best Local Similarity 44.7%; Pred. No. 0;

Matches 1168; Conservative 436; Mismatches 833; Indels 175; Gaps 36;

Qy	7	LNKISPTDRCQMTIADLOVLSFSLRKIFDDQISWGEARHLVHETIEQKNRLLEAR	66
Db	7	LLDKNRKSGDTLLITNISPMFSFEPHRTSGTLTWRETDFLYQQAQHSQKNLEELR	66
Qy	67	IFTRANPOLSGAIRLGIERDSVRSRYSDFMGARSSSFVKPGSVASPMFSAGYILTELYREA	126
Db	67	ILSRANPQANTTNINITPSTILNNSYNWFVGRHRFVKPGSTASIFSPAYILTELYREA	126
Qy	127	KOLHFSSAYHLNDRRPLADILITSQSNMDTEISTLTLSNELLELHTRKTGGSDALME	186
Db	127	KOPHPDNSOYHLNKRPPDIASLALTONNDEISTLTLSNELLIHNIOTLEKTDYNGVMK	186
Qy	187	SLSTYRQADTTPHOPYETIROVIMTHDSTLSALSRNPYMGQAEGASLILAILANISPEL	246
Db	187	MLSTYRQGTMTYPHLYFESARQAILLODKNITATSRNDDVAELMDPTSLIAIKTISPEL	246
Qy	247	YNLTIEEITEKNADALFAQNF--SENITPENFASQSWIAKYKGLSESVQKYLQMLONGYS	305
Db	247	YQLVZEEITPENSTELMKNGFTGDVL--IFKSYASLARYYDLSYDELSLFVN--LSFGKK	303
Qy	306	DSTSAYVDNISTGLVNNESKLEAKYITRVKTDYDDKKNINYFDLMYEGNQOFFIRANFKV	365
Db	304	NTNQYKNEQLITLVNDGNDTATARLIKRTKDFDYSHLNAYELIPIKENEY--KYNFSV	361
Qy	366	SREFGATL--RKNAGPSGIUCLSGPLIANTNPKSNVLSNISDSEYKNGVKIVAVRYTS	422
Db	362	KKTEPHDLFRLONGKXEYIYQDKNFVPIANTHY--SIPKILTEQTNGITLRLWRKP	419
Qy	423	STSATNQGGGIFTFESYPLTIFALKINKAIRLCITSLGSLSPNELQTVIRSDNAQGIINDSV	482
Db	420	NPSDAINANAYFKWMEFFPGDIFLLKLNKAIRLYKATGISPEDIQVIESYDVLDTISNV	479
Qy	483	LTKVFYTLFYSHRYALSFPDDAOVLNGSVINQYADDDSVSHENRLFNTPPLKGIPEADGN	542
Db	480	LGLFLFYVQYMHYNISVS DALVCHSDISQYSTKQPSHFETILFNTPLLNGOFEFSADNT	539
Qy	543	TVSIDIPIERQSTFARSALMRGLGVNSGELYOLGKLAGVLDAQNTITLSVFISSLRYLTL	602
Db	540	KLDLTGSEKNHPLYGIMKGAFRVNDTETYLTWLKLANG--GTNPFCMSIENLSILLYVRL	598
Qy	603	LARVHQLTVNELCMLYGLSPFNKKTATSLSGGELPRLVIMLYQVQTMLTBAETITTEAILW	662
Db	599	LADIHHLTVNELSMLLSVSPYNTKIALFSDTALTQLISFLFQCTQMTLLTKQMSVSDVFL	658
Qy	663	LCTPFESGNIISPEISNLNLRAPSISEDMAQSHNELOAETLAPFIATLHLASPDMARY	722
Db	659	MTTNDNTSVTLPIDENLITLTSNGULS--TUSJGDDELIRA--AAPLIHAASIQMSAKTRET	715
Qy	723	ILLWTDNLPGGDIJAGFMTLVLEKSLNANETTLQVQCHVMAQLSISVQTLRLSEABLS	782

```

QY 1787 GYSAPMDFNSALYFWELFYFTPMCFQRLLOEKQFDEATQWVINYVNPAGYIYNGHIA 1846
D 1772 NNTBPMDFNGANALYFWELFYFTPMCFQRLLOEKQFDEATQWVINYVNPAGYIYNGHIA 1831
QY 1847 PWIWCNRLPTEETSWANPLDAIDDDAVANDPMYKATFMRLLDQLILGRDMAYRELT 1906
D 1832 DYYWVRPLEEETSWANPLDSVDPAVAQHPMHYKVATFMKMLDLITRGDSAYRQLE 1891
QY 1907 RDALNEAKWYVRTLLELGDPEDEYSGQAAAPSLGSAASQTVQAAQOOLTMLGRGVVS 1966
D 1892 RDTLNEAKWYVQALTLIGDEPFSLDNDWSEPLREBAASQTMWHYQHMKLQRAAL 1951
QY 1967 KNLRTANSVLGLFLPEYNPALTDYQTLRLRLFNLRHNLSDIGQPLSLAIYAEPDTPKAL 2026
D 1952 PTKRTANSILTALFLQINKKLGQYQTLTQRLYNLRHNLTDIGQPLSLSLVATPADPSML 2011
QY 2027 LTSWQASQGSQSVLPGTILSLYRPFVMLERTENVAQLTOFCTSLISMAEHDDADELTTL 2086
D 2012 LSAAITASQGGDLPHAVPMYRPFVILENAKMGVSLIQFGNTLLSITERQDABALABI 2071
QY 2087 LLOQGMELATOSIRTOQRTVDVDDADIAVLAESRRAQNRLEKYQOQLYDDEDINHGEORAM 2146
D 2072 LQOQSELALQSIKQDKWMAEIDADKLALQESRGAQSRPDSFNTLIDEDVNAGEKQAM 2131
QY 2147 SILDDAAAGQSLAGQVLSIAEGVADLVPNVFGLAGCGSRGWAALRASVMSLSATASQVS 2206
D 2132 DYLSSSVLSTSGTALHMAAADAALVPMYGFVAGSRFGALFNASAIQIEISASATRIA 2191
QY 2207 ADKISRSAYRRRQEWIQRNADGEVQMDAQLESKIRREAAQOMQVEYQSTQOANTQ 2266
D 2192 ADKISQSIYRRRQEWIQRNABEIKQIDAQATLAVRREAAVQKYLETQOQATQ 2251
QY 2267 AQLELQKFTNKALYSWNRGKLSAIYQFPDLTQSFCLMAQELRRLTDNGVTFIRGG 2326
D 2252 AQLAFLOKFSNALYNLWLRGLSAIYQFPDLVLSCLMAEQTYQYELNNAAHFIKPG 2311
QY 2327 AWWGTTAGLMAGETLLALLAEMKVMLEDERDALEVRTVSLAQFYQALSSDNFNLTKL 2386
D 2312 AWHGTYAGLAGETLLMLAQMEKSYLEKDERALESVTRTVSLAEVYAGLTENSFTLKDKV 2371
QY 2387 TOFLREGKNGVASCNELKLSNRQTEASVRLSDLIKFSDFESIGNTRQLQKQSVTLPAL 2446
D 2372 TELVNAGSGAGTTLNGLNVEGTQLQASLKLSLDLNIATDYPDGLGNTRRIQISVTLPAL 2431
QY 2447 VGPVEDIRAVLNYGGSIVMPRCSAIALSHGVNDGSGQFMDPNDSKYLPFFEGISVNDGSG 2506
D 2432 LGPYQDVRAILSYGGSTWMPRCCKAIALSHGVNDGSGQFMDPNDKAYLPFFELPVADTGT 2491
QY 2507 LTLSPDADTRQKALLESLSLIDILHIRTYS 2538
D 2492 LTLSPFGISGKQKSLLSLIDILHIRTYS 2523

RESULT 7
AAV33729
ID AAV33729 standard; protein; 2522 AA.
XX
AC AAV33729;
XX
DT 09-NOV-1999 (first entry)
XX
DE DE
XX
XX Photorhabdus luminescens hph2-encoded insecticidal toxin.
XX
XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
XX
XX Photorhabdus luminescens.
XX
XX WO942589-A2.
XX
PD 26-AUG-1999.
XX
PP 18-FEB-1999; 99WO-EP001015.

```

```

XX 20-FEB-1998; 98US-00027080.
PR 20-JAN-1999; 99US-0116439P.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM,
PI Chen JS;
XX
DR WPI: 1999-527479/44.
DR N-PSDB; AAZ06831.
XX
PT New nucleic acid from Photorhabdus luminescens encoding insecticidal
PT toxins, used for making resistant transgenic plants.
XX
PS Claim 26; Page 133-140; 148pp; English.
XX
CC This sequence represents the hph2 gene encoded insecticidal toxin from
CC Photorhabdus luminescens. It is one of three insecticidal toxins
CC (AAV33728-Y33730) encoded by open reading frames (orfs) in a 38kb
CC fragment of P. luminescens DNA (AAZ06831). The hph2 gene was identified
CC using probe #2 which was amplified from P. luminescens genomic DNA using
CC PCR primers AAZ06829-Z06830. P. luminescens is a member of the
CC Enterobacteriaceae family and is a symbiotic bacterium of nematodes of
CC the genus Heterorhabditis. The nematodes colonise insect larvae, kill
CC them, and their offspring feed on the dead larvae. However, the
CC insecticidal agents are produced by P. luminescens rather than the
CC nematodes. The toxins have activity against Lepidopteran insects such as
CC Cabbage Looper (Trichoplusia ni), European Corn Borer (Ostrinia nubilalis)
CC and Fall Armyworm (Spodoptera frugiperda) and also against Coleopteran
CC insects (e.g., Colorado Potato Beetle, Leptinotarsa decimlineata). In
CC addition the toxins are active against strains resistant to known
CC insecticides. The DNA sequence can be used to generate transgenic plants
CC of various species that are resistant to economically important insect
CC pests and also for recombinant production of the toxins for use as
CC insecticides
XX
SQ Sequence 2522 AA;
XX
Query Match 39.5%; Score 5148.5; DB 2; Length 2522;
Best Local Similarity 43.0%; Pred. No. 0;
Matches 1121; Conservative 464; Mismatches 822; Indels 197; Gaps 55;
QY 22 LADLOYSFSLRKIFDQLSWGEARLYHETIEQKNRLLEARIFTRANPQLSGAIRL 81
D 28 LTDLSHSSFNFEHQVSEHLSWSEADLYHDAQQAQKQNRLYEARILKRTNPQLQNAVHL 87
QY 82 GIERDSVSR-SYDEMFGARSSSVKPGSVASMFSPAGYLTLYREAKDLHFPSSAYHLDN 140
D 88 AIVAPNAELIGYNNQFSGRASQYVAPGVSSMFSPAAYLTLYREARNLHASDVYRLDT 147
QY 141 RRPDLADLTLSQSNMDEITLTLNELLLEHITKCTGDS-DALWESLSTVROAIDTPY 199
D 148 RRPDLKSWALSQQNMDTELSTLSNELLLESIKTESKLDNTQVWMLSAFRPGATPY 207
QY 200 HQPYETIRQVIMTHDSTLSALSRNPEVMQAGEASLLAILANISPELYNLTETEKN 259
D 208 HDAYENVKVIQLQDPGLEQLNASPAIAGLMHQASLLGINASISPELFNLTETEKN 267
QY 260 DALFAQNPSENIPTENFASQSWIAKYGLELSEVOKYGLMQLON-GYSOSTSAVYNDISTG 318
D 268 BELYKKNFG-NIEPASLAMPYELRRYVNLSDLELSQFICKASNFQQEYSN---NQLITP 323
QY 319 LVVNNESKLEAVKIRVKTDDYDKNIYFD--LWYEGNQPFIRANFKVSREFGATLRKN 376
D 324 IVNSNDGTVKIRITR----EYTTNANQVDVELFPYGGYQINLYKFKDSRDQVYSLSK 379
QY 377 AGPSGIVGSLSGPLIANTNFKSNLSNISDSSEYKNGVKIYAYR-YTSSTSATNQGGIPT 435
D 380 LNDKRELIRIEGAPQVNIYSEHI--TLSTTDISQPFELGTRVYVPSSSWA--YAAAKFT 435
QY 436 FESYPLTITFALKLNKAIKRLCLTSGLSGPNELQTVRSNDAQIINDSVLTQVFTLYFSHR 495

```







## RESULT 8

AAW56572  
ID AAW56572 standard; protein; 2516 AA.

XX AC AAW56572;

DT 07-AUG-1998 (first entry)

XX DE Toxin TcdA, encoded by the tcdA gene from genomic region tcd.

XX KW Photorhabdus luminescens W-14; nematode; symbiotic; Heterorhabdus; tca;  
XX KW tcd; insecticidal activity; toxin; Lepidoptera; Coleoptera;  
XX KW Hymenoptera; Diptera; Dictyoptera; Acarina; Homoptera; Southern;  
XX KW Western corn rootworm; Colorado potato beetle; mealworm; boll weevil;  
XX KW turf grub; beetle armyworm; black cutworm; cabbage looper; codling moth;  
XX KW corn earworm; European corn borer; Tobacco hornworm; budworm.

XX OS Photorhabdus luminescens.

XX PN W09808932-A1.

XX PD 05-MAR-1998.

XX PF 05-MAY-1997; 97WO-US007657.

XX PR 28-AUG-1996; 96US-00705484.

XX PR 06-NOV-1996; 96US-00743699.

XX PR 06-NOV-1996; 96WO-US018003.

XX PA (DWC ) DOWELANCO.

XX PA (WISC ) WISCONSIN ALUMNI RES FOUND.

XX PI Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;

XX PI Ffrench-Constant RH, Rocheteau TA, Blackburn MB, Hey TD, Merlo DJ;

XX PI Orr GL, Roberts JL, Strickland JA, Guo L, Ciche TA, Sukhapinda K;

XX DR WPI; 1998-179427/16.

XX DR N-PSDB; AAV29928.

XX PT Isolated toxins from Photorhabdus luminescens strains - useful for

XX PT control of insect pests.

XX PS Claim 34; Page 224-231; 321pp; English.

XX CC The present sequence represents a protein named TcdA of the bacterium  
XX CC Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the  
XX CC nematodes of the Heterorhabdus genus. The bacterium has at least 4  
XX CC distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are  
XX CC produced from these regions that are associated with insecticidal  
XX CC activity. The native toxins are secreted proteins. The proteins are toxic  
XX CC to insects upon exposure and especially when ingested. The nucleic acid  
XX CC sequence can be used to produce transgenic plants, baculoviruses or  
XX CC microbial hosts for toxin production. They can be used to control insects  
XX CC pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,  
XX CC Dictyoptera, Acarina or Homoptera orders, especially the Southern or  
XX CC Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,  
XX CC turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,  
XX CC corn earworm, European corn borer or tobacco hornworm or budworm

XX SQ Sequence 2516 AA;

Query Match 38.7%; Score 5043; DB 2; Length 2516;

Best Local Similarity 42.6%; Pred. No. 0;

Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

XX 22 LADLOLSESELRKIFDDOLSWGEARHLYHETIEQKNNRLEARIFRANPQLSGAURL 81

XX 21 LTDISHSFNEFRQVSEHLSWSETHLDYHDAQQAQKNNRLEYEARILKEANPQLQNAVHL 80

XX 82 GIERDSVR-SYDEMFGARSSFVKPGSVASMPSPAGYITELYREAKDLHFSSSAYHLDN 140

Db 81 AILAPNAELIGYNNQFSGRASQYVAPGTVSSMPSPAPAYTELYREARNLHSDSVYLD 140

Qy 141 RRPDLADLTLSQSNMDEITSLTSLNELLEHI-TRKTCGSDSALMESLSTYRQAIIDTPY 199

Db 141 RRPDLKSMALSQNNMDIELSTLSNELLESIKTESKLENTKYMELMSTRPSCATPY 200

Qy 200 HOPYETIRQVIMTHDSTLSALSRNPEVMQAGASLLAILANISPELYNILEETEKNA 259

Db 201 HDAYENVREVIQLQDPGLEQLNASPAIAGLMHQASLLGINASISPELFNILEETEGNA 260

Qy 260 DALFAQNFSENITPENFASQSIKYKYGLELSEVOKYLGMLON-GYSDSTSAVVDNISTG 318

Db 261 EELYKKNFG-NIEPASLWPEYKRYNLSDBLSQFICKASNFQOQEVSN---NQLIITP 316

Qy 319 LYVNESKLEAYKTR-VKTDYDKNINYFDLMYEGNNQFFIRANPKVSRFCAT----- 372

Db 317 VVNSSDGTGVYRITREYTNAYQMDVELFP--FGEN---YRLDYKFNFNASYLSIK 371

Qy 373 -----LRKNAGPSGIVGSLGPIANTNFKSNVLSISDSEYKNGVKIYAYRYTSSTS 425

Db 372 LNDKRELVRTEGAPO-----VNIEYSANITLTAD--ISQPFELGTRVLPSGS 418

Qy 426 ATNQGSGIFTPEYPLTIPALKNKAIKRLCLTSLSPNELOTIVRSDNAQGIINDSVLTK 485

Db 419 WA-YAAAKFTVEYNQYSFLLKNKAIKRLSRATELSPITLGEIVSRVNLQDINTVLGK 477

Qy 486 VFYTLFYSHRYALSFDQAVLNGSVINQVADDSVSHFNRLFNTPPLKGIKEADGNTVS 545

Db 478 VFLTKYMQRYAHAEALILCNAPISQSYDNQPSQDFRLFNTPLLNQYFSTGDEED 537

Qy 546 IDPDEQSTFARSALMRGLGVNSBELYQLGKLAGVLDQNTTILSVFVLSLYRLTLAR 605

Db 538 LN-SGSTGDRKTIKRAFNDVDSFLRLKLTIDHNDKDGKKNLKNLSNLYIGKLLAD 596

Qy 606 VHLTVNELCMYGLSPFNGKTTAS-LASGELPRVIMLYOVYOMLTEAITEATLWLC 664

Db 597 IHLQTLSDLELL-LAVGEGKTNLSAISKQATIRKLTITSLWHTQKSVQFQFIWT 655

Qy 665 TPEFSGNISPEISNLNLRPSI---SEDMAOSHRELQAEILAPPIAATLHLASPDMA 721

Db 656 STSYNKTITPEIKNLLDTVYHGLQGFQDKAD-----LLHYMAPYIAATLQSSENV 709

Qy 722 YILLWTDNLRPGGLDIAGFMT-----LVLESINANETTO----LVQFCHVMAQLSLS 770

Db 710 SVLLWADKLQGD-----GAMTAKEFMDWMLNTKYTFEGSSEAVETQEHIVQYCALAQL 765

Qy 771 VQTLRLSRAELSVLVISGFAVLGAKNQAPAGOHNIDTFLPSLYRPHOWINGLNGPSTLDM 830

Db 766 YHSTGINENARFLFTKPK-EMFGAATGAAPAHADALSILMLTRFADWVNALGEKASVLA 824

Qy 831 LRQQLTLADRLASVNLGDISVMTQAMVSA-----GVNOLQCMODINTVLQWIDVA 880

Db 825 FEANSLTAQLADANLNDANLLQASIQAHQNHQLPPVTPPENAFSCWTSINTILQWVNA 884

Qy 881 SALHTMPSVIRTLNIRVITALKAESNLPSWDEWQTLAENWEAGLSTQAOATLADYTA 940

Db 885 QQLNVAPOGVSAVLGLDIQSMKET---PTYAOWENAAAGVLTAGLNSQANTLHAFLE 940

Qy 941 RLSSVLCHWFLANIQEGVSLHSDRLYSYFLINDQVSSAIKTTTLARAIAGIQLVYINRA 1000

Db 941 SRSAALSTYIRQVAKAAAAIKSRDDLYQLINDQVSSAIKTTIRIAIASIQLVYINRA 1000

Qy 1001 LNRIEPNARADVSTRQFFTDW--TVNNRYSTWGSVRLVYYPENYIDPTORIGOTRMMDEL 1059

Db 1001 LENVEENANSVISHQFFIDWQKYNRYSTWAGVSLVYYPENYIDPTWIRIGOTKMDAL 1060

Qy 1060 LENISQSKSLSDTVEDAFKTYLTRPETVADLVKVSAYHDVNSNTGLTFWVGQTRNLE 1119

Db 1061 LQSVSQSLNADTVEDAFMSVLTSPQVANLKVISAYHDNINNOQGLTYFGLSETDAGE 1120

Qy 1120 YWYRNVDI-SRMQAGELAAANAKWTKIDTAVNPKYDAIRPVIFRERHLIWEKEEVAK- 1178

Db 1121 YWYRVDHSHKFDGKFAANAWSEWHKIDCPINPYKSTIRPVYKSLRYLLWLEQKEITQ 1180

QY 1179 --NGTD--PVETVDRFTLKLAFLRHDSWSAPSWYDITTOVEAVTDKPKDTERLALAASG 1234  
DB 1181 TGNKSGQYQOTEDRYELKJLAHRYDGTWNTPTTFDNNKISIEL--KLEKNRAPGLYCAG 1238  
QY 1235 FQGBDILLVYVTKGYSDFGSGMKNVAGMTIYVGDGSKFCKMENTALSRYQLKNTFDII 1294  
DB 1239 YQGBDILLVYVTKGYSDFGSGMKNVAGMTIYVGDGSKFCKMENTALSRYQLKNTFDII 1294  
QY 1295 HTQGNDLVRKASYRPAQDFEVPASLNMG--AIGDSDLTVMENGINPOITSKYSDNLAI 1352  
DB 1295 DTNN--VRRVNNRYAEVIEIPSSVSRKDYWGWDYLSWVYNGDIPNTYKAASSDLKI 1351  
QY 1353 -----TLHNAAFVVRDGSNVRNKQISAMKLTGVDGSKSOYNARFIANTVHYGGY 1405  
DB 1352 YISPKLRRIHNG-----YEGQ-----KRQCNLMNKYG-----KLGDKEFIV-----Y 1388  
QY 1406 SDLG-GRITVYVTKGNY--IASVOGHLMNADYTRRLILTPVNNYARLFEF-PFSPNITL 1462  
DB 1389 TSLGWNPNNSNKLIFYVYQYSGNTSGLNQR--LLPHRTTYTPSKVEAMIPGAKSLT 1446  
QY 1463 NTVFTVGS-----NKTSDFKKCSYAVDGNNSQGFQIPSSYQSGWLDIDTGINNTDII 1515  
DB 1447 NQNAAGDDYATDSLKPDDLKQYIFMTD---SKG---TATDVSQGPVEINTAISPAKVQ 1499  
QY 1516 ITVMAGSKTHTFTASDHASLPANSPDAMPYTPFKPLEIDASSLAFTNRIAPLDIVFETKA 1575  
DB 1500 IIVKAGGKEQFTADKQVSIQPSFSDENMYQFNALIEDGSLNFINNSAIDVTFTAPA 1559  
QY 1576 KGRVLG---KIKOTLSKRVNVPEDILFLRETHSGAQYMQLVGVYRIRLNTLLASOLV 1631  
DB 1560 EDGRKLGVESPIVTLKV-----STDNALTLLHNENGAQYMQVQSYRTRLNTLFAROLV 1614  
QY 1632 SRANTGIDTILTMETQRLPEPLGSGFPANFVLPKYDPAEHDGRWFKIHXGNVCGNTGR 1691  
DB 1615 ARATGTIDTILSMETQIQEPLGKGFYATFVIPPYNLSHGDGRWFKLYKHVVVDNNS- 1673  
QY 1692 QPYISGMLSDTSETSWTLF-----VPAEYGMHEGVRLGVGYQKITYDNTWESAFFYD 1746  
DB 1674 HIIYSQGLTDTN--INITLFIPLDDVPLNQDHY-----AKYMTFTKSGSDGTWGHGFVRD 1728  
QY 1747 ETQKQFVLINDADHDSGMTQOQIVNKKYKGLNVSITATGYSAAPDFNSASALYWELF 1806  
DB 1729 D--KGIVIN-----PKSILTHFESVNVNINIS-----SEPMDFSGANSLYFWELF 1772  
QY 1807 YPTMCMCFQLQEQFDEATQWYVYVNPAGYVINGEIAPIWNCRPLEETTSWNPAL 1866  
DB 1773 YVTPMLVAQRLLEHQNFDEANRWLYVNSPSGYIVHGQIQYQWNVVRPLEETTSWNSDPL 1832  
QY 1867 DAIDPDVAQNDPMHYKIATFMRLLDQLLRGDMAYRELTRDALNEAKWYVTRILELGD 1926  
DB 1833 DSVDPDAVAQNDPMHYKYVSTFMRLLDQLLRGDMAYRELTRDALNEAKWYVTRILELGD 1892  
QY 1927 EPEDYGSQQAAPSLSGAASQVQAAQODLTMKRG--GVSXKLRTANSVLGLFLPEY 1983  
DB 1893 KPYLPSTWSDPRDLRAADITTONAHSAI VALAQNIPTPAPLSIRSANLTDLFLPQI 1952  
QY 1984 NPALTDYQOTLRRLRFLNHLNLSIDGQPLSLAIYAIEPTDPKALLTSMVOAGSGGSAVLPG 2043  
DB 1953 NEVMANYQTLAQRVYVNLRLNLSIDGQPLSLAIYAIEPTDPKALLTSMVOAGSGGSAVLPG 2012  
QY 2044 TSLYRFPVMLERTRNLVALQTOFTGTSLSMAEHDDADELTLLLOQGMELATQIRIQO 2103  
DB 2013 FMSLWRFPHMLNARGWVSQLTFQSTQNTIIRERQDAEALNALLQNAELTLTSLISQD 2072  
QY 2104 RTVDEVDADIAVLESRSQAORLEKYQOYDIEDINHGEORAMSLIDAAAGSLAGQVLS 2163  
DB 2073 KTIIELDRAEKTVEKSKAGAQSRFSDYSGKLYDENINAGENQANTLRASAGLITTAQASR 2132  
QY 2164 TAEGVADLVPNVFLGACGSRWGAAALRASAVNSISATASQVSAQYADKISRSEAYRRRREW 2223  
DB 2133 LAGAAADLVPNIFGPGAGGSRWGAAATAEATGYVMEFFSANVMNTADKISQSETYRRRREW 2192

QY 2224 EIORDNADGEVKQMDAQLESKIRREAAQMOVQEYQETQQAHTQAQLELLORKFTNKALYS 2283  
DB 2193 EIQRNNAEALKQIDAQLKSLAVREAAVLQKTSIKTQEQEQTSOLAFQKFSNALYN 2252  
QY 2284 WMRKLSAIYQFPDLTQSFCLMAQEAALRRLTONGVTFFIRGGAWNGTTAGLMAGETLLL 2343  
DB 2253 WLRGLAAIYQFPDYDLAVARCLMAEQAYRWELNDSARFIKPGAWQGTIYAGLLAGETLML 2312  
QY 2344 NLAWEKWLRLERDERALEVTRTVSLAQFYQALSSDN--FNLTEKLTOFLREGKGNVGASG 2401  
DB 2313 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKONGFSLAQAEIDKLVSOGSGSAGSGN 2372  
QY 2402 NELKL-----SNRQIEASVRLSDLKIFSDYPSLGNTRQLKQVSVTLPALVGPYEDIRAV 2456  
DB 2373 NNLAFGAGTDTKTSLOASVSFADLKIREDPASLGKIRRIKQISVTLFALLGPYQDVQAI 2432  
QY 2457 LNYGGSIVMPGCGSAIALSHGVNDGSGQFMDLDFNDSRYLPPFGISVNDGSLTSLSPDAT- 2515  
DB 2433 LSYGDKAGLANGCEALAVSHGMNDGSGQFLDFNDGKFLPFEGIAI-DQGTTLTSLSPNASM 2491  
QY 2516 ---DROKALLESLSDIILHRYTIR 2537  
DB 2492 PEKQQAATMLKTLNDIILHRYTIR 2516

## RESULT 9

AAB72609  
ID AAB72609 standard; protein; 2516 AA.  
XX AAB72609;  
XX 04-MAY-2001 (first entry)  
XX Photorhabdus tcda toxin.  
XX Tcda; Tcda; insect toxin; plant; insect resistance.  
XX Photorhabdus sp.  
XX WO200111029-A1.  
XX 15-FEB-2001.  
XX 11-AUG-2000; 2000WO-US022237.  
XX 11-AUG-1999; 99US-0148356P.  
XX (DWC ) DOW AGROSCIENCES LLC.  
XX Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;  
XX Sukhapinda K, Merlo AO;  
XX WPI; 2001-191536/19.  
XX N-PSDB; AAF58778.  
XX Novel polynucleotide sequence encoding insect toxins, useful for  
XX producing transgenic plants having resistance to insects, especially corn  
XX rootworm.  
XX Disclosure; Page 51-61; 106pp; English.  
XX The present invention provides the protein and coding sequences of  
XX modified versions of the photorhabdus Tcda and Tcda toxins. These are  
XX suitable for expression in plants. The toxins are effective against  
XX insects upon ingestion, and the sequences provided can be used to produce  
XX transgenic plants with insect resistance. The present sequence is the  
XX photorhabdus Tcda protein  
XX Sequence 2516 AA;

Query Match 38.7%; Score 5043; DB 4; Length 2516;  
Best Local Similarity 42.6%; Pred. No. 0;  
Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;







XX (UYBA-) UNIV BATH.  
 XX Ffrench-Constant RH, Waterfield NR;  
 XX WPI: 2004-411735/38.  
 XX N-PSDB; ADP18613.  
 XX New isolated Photorhabdus luminescens nucleic acids, useful for  
 PT expressing orally active insect toxin or for generating transgenic plants  
 PT with enhanced resistance to insects.  
 XX Disclosure; SEQ ID NO 2; 118pp; English.  
 XX The present invention describes DNA sequences from the tcd genomic region  
 CC of Photorhabdus luminescens W-14. Also described: (1) a transgenic  
 CC monocot or dicot cell having a genome comprising a nucleic acid sequence  
 CC that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant  
 CC with a genome comprising a nucleic acid sequence that encodes SEQ ID  
 CC NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a  
 CC progeny of seed; (5) a method of producing Toxin A of Photorhabdus  
 CC luminescens W-14 in a heterologous host; and (6) a method of producing an  
 CC orally active insect toxin. The nucleotide sequences are useful for  
 CC heterologous expression of orally active insect toxin. They can also be  
 CC used for generating transgenic plants with enhanced resistance to  
 CC insects. The present sequence represents Photorhabdus luminescens W-14  
 CC tcdA1, which is given in the exemplification of the present invention.  
 XX Sequence 2516 AA;  
 Query Match 38.7%; Score 5043; DB 8; Length 2516;  
 Best Local Similarity 42.6%; Pred. No. 0;  
 Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;  
 22 LADLOVLSFSELRKIFDDOLSGEAREHLYHETIEQKNNRLEARIFRANPQLSGAIRL 81  
 21 LTDISHSEFNEFRQVSEHLSWSETHLDYHDAQQAQKQNRLEYEARILKRNAPQLNAVHL 80  
 82 GIERDSVR-SYDENMFGARSSFFVPGSVASMPSPAGYLTLYERAKDLHFSSSAYHLDN 140  
 81 AILAPNAELIGYNNQFSGRASQYVAPGTVSSMFPSPAAVLTLYERARNLHASDSVYLD 140  
 141 RPDLDADLTLSQNMDETSTLTLSNELLLEHI-TRKTGSDALMESLSTVROAIDTPY 199  
 141 RPDLSKWSALSQNMDELTLSNELLLEHI-TRKTGSDALMESLSTVROAIDTPY 200  
 200 HOPYETIRQVIMTHDSTLSALSRNPEVMQAGASLAILANISPELYNLTETEETKNA 259  
 201 HDAYENVREVIQDPGLEQLNASPAIAGLMHQASLILGINASISPELFNLTETEETGNA 260  
 260 DALFRONTSENITPENFASQSWIAKYGLESEVOKYGLMLQON-GYSDSTSYVDNISTG 318  
 261 EELYKKNFG-NIEPASLAMPYELKRYNLSDELSQFTGKSNFQOQBYSN---NQLITP 316  
 319 LVVNNEKLEAYKTR-VKTDDYDKNINFDLMYGNNOFFFRANFKVSRFEGAT- 372  
 317 VNNSDGTGVKVRITREYTTNAYQMDVELFP--FGEN---YRLDYKKNFNASYLSIK 371  
 373 -----LRKNAGPSGIVGSLGPIANTNFKSNLSNISDSYKNGKVIAYRYTSSTS 425  
 372 LNDKRELVRTEGAPO-----VNIEYSANITLTAD--ISQPFELGLTRVLPSSG 418  
 426 ATNQGGLITFTFESYPLTTPALKLNKAIRCLTSGLSPELQTVIRSDNAQGLINDSVLTK 485  
 419 WA-YAAAKFTVEYNQYSPFLKLNKAIRLSRATLSPTILEGIVRSVNLQDINDTVLKG 477  
 486 VPYTLFYSRYALSDDDAOVLNGSVINOVADDSDVSHFNRLPENTPLKGIPEADGNTVS 545  
 478 VFLTKYMQRYAIHAETAILICNAPISQBSYDNQPSQDFRLENTPLANGQYFESTGDEID 537  
 546 IDPDEQSTFARSALMRGLGVNSGELYQGLAGVLDAGNTITLSVVFSSLYRUTLLAR 605  
 538 LN-SGSTGDRKTLKRAFNIDVSLFRLLKLTIDHNKDGKIKNNLKNLSNLYIGKLLAD 596

QY 606 VHQLTVNELCWLYGLSPNGKTTAS-LSSGELPRLVIMLYOVTOMLTEABITTEAIIWLIC 664  
 DB 597 IHQUTIDELDLLL-IAVGEKTNLSAIDSKOLATLIRKLTNTITSWLHTQKWSVQLFTMT 655  
 QY 665 TPEFSGNISPEISNLLNNLRPSI---SEDMAQSHNRELQABILAPFAATLHLASPDMA 721  
 DB 656 STSYNKTLTPEIKNLLDTVYHGLQGFQKAD-----LLHVMAPYIAATLQLSSENA 709  
 QY 722 YLLWNTNLRPGGLDIDAGFWT-----LVUKESLNANETQ-----LVQFCHVMAQLSLS 770  
 DB 710 SVLLWADKLPQGD---GAMTAERFMDWLNKTYTPGSEAVETQEHIVQYCOALAQLSMV 765  
 QY 771 VQTLRLSEAEISLVISGFAVLGAKNOPAGOHNTDTLPSLRFHQWINGLNCNPGSDTUDM 830  
 DB 766 YHSTGINEANRLFTVTP-EMFGAATGAPAHADALSILMLTRFADWVNALKEKASSVLAA 824  
 QY 831 LRQOULTADRILASVMGLDISMVTQAMVSA-----GVNQLQCWODINTVLIQWIDVA 880  
 DB 825 FEANSLTAEQLADAMNLDANLLQASIQAOHQHLPVPTPENAFSCWTSINTILQWVNA 884  
 QY 881 SALHTMSVIRTLNIRYVTALNKAESNLPSEWQTLAENWEAGLSLTOOQATLADYTA 940  
 DB 885 QQLAVAPQGVSGALVGLDYIOSMKET---PTYAQWENAAAGVLTAGLNSQQANTLHAFDE 940  
 QY 941 RLSSVLCNWFILANIOPEGVSLHSRDDLYSYFLIDNOVSSAIKTTRLAEAIAGIOLYINRA 1000  
 DB 941 SRSAALSTYIRQVAKAAAALIKSRDDLYQYLLIDNOVSAALKTTTRIAEAIASIOLYINRA 1000  
 QY 1001 LNRIEPNARADVSTRQFFTDW-TVNNRYSTWGGVSRVLYYPENYIDTPQIRGOTRMDEL 1059  
 DB 1001 LENVEENANSVSIHQPFIDWDKYNKRYSTWAGVSQVLYYPENYIDTPMTRIGQTKMDAL 1060  
 QY 1060 LENISQSKLSRDTVEDAPKTYLTRPETVADLKVSAYHDVNNSVNTGLTWFVQOTRENLE 1119  
 DB 1061 LQSVSQSLNADTVEDAFMSYLTSEFQVANLKVISAYHDNINNOQGLTYFTGLSETDAGE 1120  
 QY 1120 YYWRNVDISRMQAGELAANAWEKTKIDTAVNPKDAIRPVIFFERLHLIWEKEEVAK- 1178  
 DB 1121 YWKSVDHUSKENDGKFAANASEWHKIDCPINPKYSTIRPVIYKSRLLYLLWLEKEIKTQ 1180  
 QY 1179 --NGTD--PVETYDRFTLKLAFLRHGDSWSPWSDYITTOVEAVTDKPKDTERLALASG 1234  
 DB 1181 TGNKSGQYQETDVRVELKLAHRYDGTWNTPTITFDVNNKISL--KLEKNRAPGLYCAG 1238  
 QY 1235 FQGEDTLVYFKYTKGSYSDGSGSNKVNAGMTIYDGSFKKMENTALSRYQLKNTFDI 1294  
 DB 1239 YQGEDTLVYFMFYNOQDTLDSY--KNASMOGLYIFADMAKMDTPQSNVYRD--NSYQOF 1294  
 QY 1295 HTQGNDLVRKASVYRPAQDFEVPASLNMGSS--AIGDLSLTVMENGNIPOITSYSSDNLAI 1352  
 DB 1295 DTNN---VRRVNNRYAEDEYIEIPSSVSRKDYGWGDYLSVMVYNGDIPINYKAAASDLKI 1351  
 QY 1353 -----TLHNAAFVRYDGSNGVIRNKQISAMKLTGVGDGKSQYGNAFIANTVRYGY 1405  
 DB 1352 YISPKLRIIHNG-----YEQQ---KRQCNLMNKYG-----KLGDKEIV-----Y 1388  
 QY 1406 SDLG-GPITVYNTKQY--IASVQGHLMADYTRRLILTPVENNYVYARLFE-PPSPNTIL 1462  
 DB 1389 TSLGVNPNNSKLMFYVYQYSGNTSGLNQGR--LLFHRDTPPYPSKVEAMIPGAKRSLT 1446  
 QY 1463 NTVFTVGS-----NKTSDFKCSYAVDGNNSQGFQIFSSYQSSGWLDDIDTGINTTIDK 1515  
 DB 1447 NONAIGDDYATDSLKPDDLKQYIFMTD---SKG---TATDVSQVPEINTAISPAKVQ 1499  
 QY 1516 ITVMAGSKTHFTTASDHISLIPANSFDMPTFKPDLIDASSLAFATNNIAPLDIVFETKA 1575  
 DB 1500 IIVRAGGKEQFTADKQVSIQPSFSDENMYQFNALIEDGSGLNFINNSASIDVTFTAPA 1559  
 QY 1576 KGRVILG---KIKOTLSVKRVNYPEDILFLRETHSGAQYMLGVYRIRLNTLLASOLV 1631  
 DB 1560 EDGRKLGYESFSPVTLTKV-----STDNALTLHNENGAQIMQWOSYTRRLNTLFRQJLV 1614

Qy	1632	SRANTGIDTILTMETQRLPPPLPGEQFPANFVLPKYDPAEAGHDERWPKIIHGNVGGN	1691
Dy	1615	ARATYTGIDTILSMETQIQEPLQKGFYATFVIPPNVLSHTGGDERWPKLYIKHVVDNNS-	1673
Qy	1692	QPYYSGLMSTSETSMTLF-----VPAEAGYVMEHGVRLGVGVOKIYDNTWESARFPD	1746
Dy	1674	HIYSGQUTDN-INITLFLPDLDDVPLNQDTH-----AKVTMTFKSPSDGTWGPHPVRD	1728
Qy	1747	ETKQOQFVLINDADHSGMTQOGIVKNIKCYKGFGLNVSIATGYSAIPMDFNASALYIWELF	1806
Dy	1729	D--KGIVTN-----PKSILTHFESVNVLNIS-----SEPMDFSGANSILFYWELF	1772
Qy	1807	YYPWMCFORLLQKQKQDEATQWIVNYNPAGYIVNGEIAPIWVNCRPLBETTSWNAIPL	1866
Dy	1773	YYPMLVAORLLHHEQNFDEANRWLKYVWSPSGYIVHGOIQNYQWNVPLLEDTSWNSDPL	1832
Qy	1867	DAIDPDVAQNDPWHYKIATFMRLLDQLILRGDMAYRELTDRALNEAKMYVTRTELLGD	1926
Dy	1833	DSVDPDAVAQHPWHYKVSTFMRTLDLLIARGDHAYRQLERDTYNEAKMYMQLHLLGD	1892
Qy	1927	EPEDYSGQWAAAPSLSGAASQTVQAAYQODLTMLGRG---GVSKNLRKNTANSVLGLFPEY	1983
Dy	1893	KPYLPLTSTWSDPRLDRAADITTCNAHDSAIVALRQNIPTPAPLSLRSANTLTDLFLPQI	1952
Qy	1984	NPALTDYWTQRLRLFLNLRNLSIDGQPLSLAIYAEPDPAKALITSMVQASQGSVAIVPG	2043
Dy	1953	NEVMNNTWQTLAQRVYNLRNLSIDGQPLPLPIYATPADPKALLSAAVATSGGKGLPES	2012
Qy	2044	TLSLYRPPVMLERTNLVAQLTQGTCTILLSMAEHDDADELTITLLQCGMELATQSTRIQQ	2103
Dy	2013	FMSLWRPPHLENARGMVSQLTQFGSTLQNIIERQDAEALNALLQNAQLIITNLSIQD	2072
Qy	2104	RTVDEVDAIVLAESRRSAQNRLBKQOQLYDEIDNHGEORAMSLIDAAAQSGIAGQVLS	2163
Dy	2073	KTIIELDABKTVLEKSKAGAQSPDSYGLKYDENINAGENQAMTLRASAAGLTVAQASR	2132
Qy	2164	IAEGVADLVPNVFGIACGGRSGWGAALPASVMSLSATASQYSGADKISRSEAYRRRQEW	2223
Dy	2133	LAGAAADLVPMIIFGFAGGSGRWGAIAEATGVVMEFSANVMNTEADKISQSETVYRRRQEW	2192
Qy	2224	EIQEDNADGEVQWDAQLESIKTIRREAAQOVVEYQFTQQAHTQAQLLQKQKTNKALYS	2283
Dy	2193	EIQRNNAEASLKTQIDAQLKSLAVRREAVALQKTSKTKQEQTOSQLAFQKQFSGNALYN	2252
Qy	2284	WMRGKLSAIYYQFPDLTQSFCLMAQBAALRRELTDTNGVTFIRGGAWNGTTAGLMAGETLLI	2343
Dy	2253	WLGRLANIYFQYDVLAVARCLMAEQYRWELNDDNSARFIKPAWQGTQYAGLAGETMLL	2312
Qy	2344	NLAEMEKVWLDERALAEVTRTVSLAQYQALSSDN--FNLTKLTQFLREGKGNVGAAG	2401
Dy	2313	SLAQMEDAHLKDKIKRALEVERTVSLAEVYAGLPKDNGPPFSLAQEIDKLVSGQSGSAGSN	2372
Qy	2402	NELKL-----SNRQIEASVRLSDLKIPSDYPESIGNTRQLKQVSVTLPALVGPYEDIRAV	2456
Dy	2373	NNLAFAGAGTDKTSLQASVSFADLKIREDPASLGKIRRTIKQISVTLPALVGPYQVQAI	2432
Qy	2457	LNYGGSIVMPRGCSAIALSHGVNDSGQFMLEDNDSRYLPEGIVSNDSGSLTSLFFPDAT-	2515
Dy	2433	LSYGDKAGLANGCEALAVSHGMDSGQQLDFDNDGKFLPPEGIAI-DOGTILTSLFFPNASM	2491
Qy	2516	---DRQKALLESIDTILHITYTIR 2537	
Dy	2492	PEKQKQATMLKTLNDIILHITYTIK 2516	
RESULT 12			
ID	ADR21575		
XX	ADR21575	standard; protein; 2516 AA.	
AC	ADR21575;		
XX			
DT	04-NOV-2004	(first entry)	
XX			







Db 2373 NNLAFCAGTDTKTSLOASVSFADLKIRIDYPASLGKIRRIKQISVTLFALLGPGYQDVQAI 2432  
 QY 2457 LNYGGSIVMPGCSAIALSHGVNDGQFMDLNDPNDKRYLPFFEGISVNDGSLTSPDAD- 2515  
 Db 2433 LSYGDKAGLANCEALANSHGVNDGQFMDLNDKRYLPFFEGIAI-DQGTLSLSPNASM 2491  
 QY 2516 ---DRQKALLESLSIIHRYTIR 2537  
 Db 2492 PEKGQATWLTNDIILHRYTIK 2516

RESULT 13  
 ADZ72157  
 ID ADZ72157 standard; protein; 2516 AA.

XX AC ADZ72157;  
 XX DT 14-JUL-2005 (first entry)  
 XX DE P. luminescens toxin A protein SEQ ID NO:4.  
 XX KW transgenic plant; vector; pesticide resistance; insecticide resistance;  
 XX KW herbicide resistance; insect resistance; toxin A.  
 XX OS Photorhabdus luminescens.

XX US2005102713-A1.  
 XX PD 12-MAY-2005.  
 XX PF 07-NOV-2003; 2003US-00703280.  
 XX PR 07-NOV-2003; 2003US-00703280.

XX PA (MERL/) MERLO D J.  
 XX PA (LIUD/) LIU D.  
 XX PA (BURT/) BURTON S L.  
 XX PA (GLAN/) GLANCY T P.

XX PI Merlo DJ, Liu D, Burton SL, Glancy TP;  
 XX WPI; 2005-384062/39.

XX DR New nucleic acid construct comprises a structural gene of interest  
 XX PT functionally linked to untranslated regions (UTR), useful for enhancing  
 XX PT expression of a recombinant nucleic acid sequence in transgenic plants  
 XX PT and plant tissues.

XX PS Example; SEQ ID NO 4; 99pp; English.

XX CC The invention relates to a novel nucleic acid construct comprising at  
 CC least one structural gene of interest functionally linked to one or more  
 CC of an untranslated region (UTR) selected from 3 fully defined 40-260  
 CC base pair (bp) sequences (ADZ72154, ADZ72155 or ADZ72156), given in the  
 CC specification or its derivatives or functional equivalents. Also claimed  
 CC are a plant or plant cell transformed with the nucleic acid construct, a  
 CC vector comprising the nucleic acid construct, a method for recombinantly  
 CC producing a peptide or protein, a method of increasing expression of a  
 CC gene in a plant cell, a recombinant DNA construct comprising a gene of  
 CC interest and at least of a UTR selected from a 5' or a 3' UTR from an  
 CC osmotin gene, a 5' UTR from an osmotin gene or its derivative  
 CC functionally linked to a heterologous gene, a 3' UTR from an osmotin gene  
 CC or its derivative functionally linked to a heterologous gene, and a  
 CC recombinant nucleic acid construct comprising 5' and/or 3' UTR isolated  
 CC or derived from an osmotin gene, the 5' and/or 3' UTR functionally linked  
 CC to a structural gene of interest. In the nucleic acid construct, at least  
 CC one structural gene of interest comprises a gene capable of conferring a  
 CC non-native phenotype in a plant or an insecticide or herbicide resistance  
 CC in a plant. At least one structural gene of interest also comprises an  
 CC insect resistance gene isolated or derived from 5 fully defined 1750-7302  
 CC bp sequences (ADZ72159-ADZ72163), given in the specification. At least  
 CC one structural gene of interest further comprises a fully defined 7560 bp

CC sequence (ADZ72158), given in the specification or its derivative. The  
 CC nucleic acid construct is useful for enhancing expression of a  
 CC recombinant nucleic acid sequence in transgenic plants and plant  
 CC tissues, or organisms using 5' or 3' UTR regions isolated or derived from  
 CC a tobacco osmotin gene. The present sequence represents the P.  
 CC luminescens toxin A protein, an insecticidal protein used in a vector of  
 CC the invention.

XX SQ Sequence 2516 AA;

Query Match 38.7%; Score 5043; DB 9; Length 2516;  
 Best Local Similarity 42.6%; Pred. No. 0;  
 Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;  
 QY 22 LADQLYLSFSELRLKIFDDQLSWGEARHLYTHETIOKNNRLLEARIFRANPQLSGARL 81  
 Db 21 LTDISHSFNEFRQVSEHLSWETHDYHDAQQAQKONRLYEALIKRANPQLQNAVHL 80  
 QY 82 GIERDSVSR-SYDEMFGARSSSVKPGSVAMFSPAGYLTLYTEAKDLHFPSSAYHLDN 140  
 Db 81 AILAPNAELIGVNNQFSGRASQYVAPGTVSSMFPSPAYLTLYTEAKDLHFPSSAYHLDN 140  
 QY 141 RPDLDADLTLSQSNMDEITLTLNELLLEHI-TRKTGGSDALMESISTYRQADITPY 199  
 Db 141 RPDLSKMSALSOQNMDEITLTLNELLLESIKTESKLENTKYVMELSTFRPGATPY 200  
 QY 200 HOPYETIRQVIMTHDSTLSALSRNPEVVMQEGASLLAILANISPELYNLTETEEKNA 259  
 Db 201 HDAYENREVILQDPGLEQLNASPAIAGLMHQASLLGINASISPELFNLTETEEGNA 260  
 QY 260 DALPAQNFSENITPENFASQSWIAKYGLEISEVQKYLGMQLON-GYSOSTSAIVDNISGT 318  
 Db 261 EELYKNFG-NIEPASLAMPYELKRYNLSDEELSQFICKASNFQQQESYN---NQLITP 316  
 QY 319 LVNNESKLEAYKITR-VKTDYDKNINYFDLMYEGNNOFFFRANPKVSRFEGAT- 372  
 Db 317 VNSSDGTGVKYRITREYTNAYQMDVELFP--FGGEN---YRLDYKFNFNASYLSIK 371  
 QY 373 -----LRKNAGPSGIVGSLGPLIANTFNFKSNLYNISDSEYKNGVKIYARYTSSTS 425  
 Db 372 LNDKRELVRTEGAQO-----VNIEYSANITLTAD--ISQPEIGLTRVLPSGS 418  
 QY 426 ATNCGGIFTPESEPLTIFALKKNAIRLCITGSLSPNELOTIVRSDNAQGIINOSVLT 485  
 Db 419 WA-YAAAKFTVEEYNQYSFLKNAIRLSRATLSPTILEGIVRSVNLQDLINTDVLGK 477  
 QY 486 VFTYLFYSHRVALSFDDAQLNGSVINQYADDSVSHFNRLPNTPLKGIPEADGNTVS 545  
 Db 478 VFLTKYMQRYAIHAETALILCNAPISQSRSDNQPQDFRULFNTPLNGQYFSTGDEED 537  
 QY 546 IDPDEEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFISSLYRLTLAR 605  
 Db 538 LN-SGSTGDWKTILKRAFNDIDVSLFRLLKAITDHDNDKDKGKKNLKNLSNLYIGKLLAD 596  
 QY 606 VHQLTVNCLMYGLSPFNGKTTAS-LSSGBELPRVIMLYQVQMTLTRAETITEAILWLC 664  
 Db 597 IHQTLIDELDLLL-IAVGEKTNLSAISDKQLATLIRKLTNTITTSWLTKQKWSVOLFIMT 655  
 QY 665 TPESGNTSPISNLLNLRPSI-----SDMAQSHNRLEQAEILAPPIAATLHASPDMAR 721  
 Db 656 STSNKTLTPRIKLLDVTYHGLGQDFDKDAD-----LLHVMAPYIATLQLSSENVAH 709  
 QY 722 YILLWTDNLRPGGLDIAGFWT-----LVLKESLNANETIQ-----LVQFCHVMAQLSL 770  
 Db 710 SVLLWADKLQGD-----GAMTAERFWDWNTKYTPGSSAEVETQEHIVQYCOALQLENV 765  
 QY 771 VQTLRLSBAELSVLVISGFVAVLGAKNPAGQHNIDTLFSLYRHFHOMINGLNGPSSDITLDM 830  
 Db 766 YHSTGINENAFRLFVTKP-EMFGAATGAAPAHADALSILMLTRFADWNVNAGKASSVLAA 824  
 QY 831 LRQQTILADRLASVNGLDISWTOAMVSA-----GVNQLQCQWDINTVLQWIDVA 880

Db 825 FEANSLTAEQLADAMNLDANLLQASIQAOHQHLPVPTPENAFSCWTSINTILQWNVVA 884  
Qy 881 SALHTMPSVIRTLNIRVYTAALNKAESNLPSWDEWOTLAENWEAGLSLTOQAQTADYTAE 940  
Db 885 QQLNVAPOGVALVGLDYIQSMKET-PTYAOWENAAAGVLTAGLNSQQANTLHFAFLDE 940  
Qy 941 RLSSVLCNWFELANTIQPEGVSLHSDRLYSYFLIDNOVSSAJKTRTLAEAIAGIOLYINRA 1000  
Db 941 SRSAALSTYIYIROVAKAAAKSRDLDLYLLIDNOVSSAAIKTRIAEAIASIQLYYNRA 1000  
Qy 1001 LNRPEPNARADVSTROFFTDW-TVNNRYSTWGGVSRLYYPENYIDPTQRIQOTRMMDL 1059  
Db 1001 LENVEENANGSVISRFQFIDWDKYNKRYSTWAGVSQLYYPENYIDPTWIRIGQTKMDAL 1060  
Qy 1060 LENISQSKSLSDTVEDAFKTYLTRPETVADILKVSAYHDNVNSNGLTWFGQOTRENLP 1119  
Db 1061 LQSVSQSLNADPTVEDAFMSYLTSEFQVANLKVTSAYHDNINNOQGLTYFTGLSETDAGE 1120  
Qy 1120 YYWRNVDIRMQAGELAANAKWTKIDTAVNPYKDAIRPVI FRERLHLIWWKEEVAK- 1178  
Db 1121 YYWRSVDHSHKFNDDGKFAANANSEMHKIDCPINPKYSTIRPIVYKSLRLLMLWLBQKEITQ 1180  
Qy 1179 --NGTD--PVETYDRFTLKLAFRLHDSWSPSYDITTOVEAVTDRKXPDTERLALASG 1234  
Db 1181 TGNKDGQYQTFDYRYELKLAHIRYDGTWNTPIITFDVNNKISL--KLEKNRAPGLYCAG 1238  
Qy 1235 FQGEDTLVYVYTKGKSYDFGGSNKNAVAGMTIYGDGSKFOMENTALSRYSQLKNTFDII 1294  
Db 1239 YQGEDTLVMPYNOQDTLDSY--KNASMOGLYIFADMASKOMTPEQSNVYRD--NSYQOF 1294  
Qy 1295 HTQGNDLVRKASYRPAODFEVPASLANMGS--AIGDLSLTVWENGNIPOITSKYSNDLAI 1352  
Db 1295 DTNN---VRRVNNRYAEIYELPSSVSRKDYGWGDYLLSWYNGDIPINTYKAASSDLKI 1351  
Qy 1353 -----TLHNAFTVRYDGSNGVIRNRKQISAMKLTGVDGKSQYGNAFIIANTVKHYGGY 1405  
Db 1352 YISPKLRIIHNG-----YEGQ---KRQCNLNKKYG-----KLGDKEIV-----Y 1388  
Qy 1406 SDLG-GPITVYNTKYN--IASVOGHLNADYTRRLILTPVNNYARLPF-PSPNTIL 1462  
Db 1389 TSLGVNPNSSNKLMPYVYQSGNTGLNQR--LLPHRDTTPYPSKVEAWIPGAKRSLT 1446  
Qy 1463 NTVTFVGS-----NKTSDFKKCSYAVDGNNSQGFQIFSSVQSGWLDIDTGINNTDIK 1515  
Db 1447 NONRAIGDDYATDSLUNKEDDLKQYIFMTD---SKG-----TADVSGPVEINTALSPAKVQ 1499  
Qy 1516 ITVMAGSKTHFTASDHIALSPANSFDAMPYTFKPLEIDASSLAFTNNIAPLDIVPETKA 1575  
Db 1500 IIVKAGGKEQFTADKDVSIQPSPSFDERMNYQFNALEIDGSLNFINNSASIDVTFTAFA 1559  
Qy 1576 KGRVILG-----KIKOTLSVKRVNVPEDILFIRETHSQAQYMQLGVRIRLNTLASOLV 1631  
Db 1560 EDGRKLGYESFSIPVTLKV-----STDNALTLHHNENGAQYMQWQSYRTRLNTLFAROLV 1614  
Qy 1632 SRANTGIDTILTMETQRLPEPLGGRFFANFVLPKYDPAEHGDERWPKIHIHGNVGNTR 1691  
Db 1615 ARATTGIDTILISMETQNLQBPOLGKGFYATVPYPPNLSHGDERNFKLYKHVVDNNS- 1673  
Qy 1692 QPYTSGMLSDTSETSMTLF-----VPIAEGYMHGVRGLGVQYKITYDNTWESAFFYD 1746  
Db 1674 HIIYSGQLTDIN-INITLFIPLDDVPLNQDYH---AKVYMTFKKSPSDGTWGWGPHFVRD 1728  
Qy 1747 ETKOQFVLINADHDSGMTQGOIVKNIKKYKGFNLVSIATGYSAPMDFNASALYYWELF 1806  
Db 1729 D--KGIVTIN-----PKSILTHPESVNVLNINIS-----SEPMDFSGANSLYFWELF 1772  
Qy 1807 YITPMWCFORLLOEKQFDEATQWYNVYNPAGYIVNGBIAPWNCRPLEETTSNANPL 1866  
Db 1773 YITPMVLAQRLHHEQNFDEARNLKWYVWSPSGYIVHGQIQNYQMNVRPLEEDTSWNSDL 1832  
Qy 1867 DAIDPDVAQNDDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVTRLELGD 1926  
Db 1833 DSVDPDAVAQDPMHYKYSTFWRTLDDLLIARGDHAYRQLERDTLNEAKWYMQALHLLGD 1892

Qy 1927 EPEDYGSQQAAPSLSGAASQTVQAAQQODLTMLGRG---GVSKNLRTANSLVGLFLPEY 1983  
Db 1993 KPYPPLSTWSDPRLDRAADITTONAHDSAIVALRQNIPTPAPLSLSANTLDTLFLPQI 1952  
Qy 1984 NPALTDYMQTLRLRLFNLRHNLSDIGQPLSLAIYAEPDTPKALTSVMQASQGSASVLP 2043  
Db 1953 NEVMNMYQTLAQRVYNLRHNLSDIGQPLYLPIYATPADPKALLSAAVATSGGKLPES 2012  
Qy 2044 TLSLYRPPVMLERTNVLVAQLTQGTSLLSMAEHDDADELTLLLOQGMELATOSIRQQ 2103  
Db 2013 FMSLWRFPHMLNARGMVSQLTQFGSTLQNIIERQDAEALNALLQNAEAILNLISQD 2072  
Qy 2104 RTVDEVDADIAVLAESRRAQNRLEKYQOLYDEIDINEHQEORAMSLDAAAGSOLAGVLS 2163  
Db 2073 KTIIBELDAEKTVLKSKAGASRFDSYKGLYDENINAGENQAMTLRASAGLITAVQASR 2132  
Qy 2164 IAEGVADILVNPVFLGACGSRGWAALRASASVMSLSATASQYSADKISRSBAYRRRQEW 2223  
Db 2133 LAGAAADLVNPIFGFAGGSRWGAIAEATGYVMEFSANVMNTEADKISQSETYRRRRQEW 2192  
Qy 2224 EIQRDNADGEVKQMDAQLESKIRREAAQMVQYEQTOQAHTQAOLELLQKRFNKALYS 2283  
Db 2193 EIQRNNAEELKQIDAQKSLAVREARVLOKTSKTQEQTSQALFQKRFNSQALYN 2252  
Qy 2284 WMRGKLSAIYYQFDTLQSFCLMAQEALRRELDTONGVTIFIRGGAWNGTTAGLMAGETLL 2343  
Db 2253 WLRGLAAIYQFDTLAVARCLMAEQAYRWELNDDSRFIPKGAWQGTYAGLLAGETLML 2312  
Qy 2344 NLAEMEKVLERDERALEVTRTVSLAQFYQALSSDN--FNLTEKLTQFLREGKGNVGSAG 2401  
Db 2313 SLAQMEDAHLKRDKRALLEVERTVSLAEVYAGLPKDNNGPFLSAQEBIDKLVSQSGSAGSN 2372  
Qy 2402 NELKL-----SNRQIBASVRLSDLKIFSDYPESLGNTRQLKOVSVTLPALVGPYEDIRAV 2456  
Db 2373 NNLAFGAGTDTKTSLQASVSPADIKIREDPASLGRIRKIQISVTLPALGYPQDVOAI 2432  
Qy 2457 LNYGGSIVMPRGCSAIALSHGVNDSGOFMLDNDNRYLPPGIGSVNDSGSLTLGFPDAT- 2515  
Db 2433 LSYGKGAGLNGCBALAVSHGVNDSGQFQDFNDGKFLPFEGIAI-DOGTTLTLSPFNASM 2491  
Qy 2516 ---DRQKALLESLDIIILHRYTIR 2537  
Db 2492 PEKKGQATMLKTLNDIILHRYTIK 2516

## RESULT 14

AAAB72611  
ID AAB72611 standard; protein; 2517 AA.  
XX AAB72611;  
AC AAB72611;  
XX 04-MAY-2001 (first entry)  
DT Modified Photorhabdus tcdA toxin.  
XX TcdA; TcdA; insect toxin; plant; insect resistance.  
DE Photorhabdus sp.  
OS Synthetic.  
XX WO200111029-A1.  
PN 15-FEB-2001.  
PD 11-AUG-2000; 2000WO-US022237.  
PF 11-AUG-1999; 99US-0148356P.  
PR (DOWC ) DOW AGROSCIENCES LLC.  
PA Petell JK, Merlo DV, Herman RA, Roberts JL, Guo L, Schafer BW;  
PI Sukhapinda K, Merlo AO;





Db 559 LN-SGSGDWRTKILKRAFNIDDDVLSFLRLKLTIDHNDKDGKIKNNLKNLSNLIYIGKLLAD 617  
Qy 606 VHQLTVNELCMLYGLSPFNKKTAS--LSSGELPRVIMVLYQVOTMLTAEIITEAIWLIC 664  
Db 618 IHQILIDELDLL--TAVGBGKTNLSAISDKQLATLIRKLTITSLMLTQKWSVFLFMT 676  
Qy 665 TPEFGNITSPEISNLLNLRPSI-----SEMAOSHREIQAEILAPFIANTLHLPDMAR 721  
Db 677 STSYNKLTPETIKNLLDTPYHGLQDFDKAD-----LLHVMAPYIAATLQLSSENVAH 730  
Qy 722 YLLMTDNLPRGGLDIAGPMT-----LVLAESLNANETQ-----LVQFCHVMAQLSLS 770  
Db 731 SVLLWADKLQPD-----GAWTAEKFWMDLNTYTFGSSSEAVTQHEIYOICQLAQLEKV 786  
Qy 771 VOTLRSEAEBSLVVISGPAVLGAKNQAGQHNIDTFLSLYFHOWINGLNGPFGSDTIDM 830  
Db 787 YHSTGINENAPRLFVTKP--EMFGAATGAAPAHADALSILMLTRFADWVNALGKASVLAA 845  
Qy 831 LRQOITLADRSLASVGLDISMTQAMVSA-----GVNLOCWQODINTVLOWIDVA 880  
Db 846 PEANSULTAEQLADAMNLDANLLLOASIOAQNHOHLPPVTPPENAFSCWTSINTILOQWVA 905  
Qy 881 SALHTMPSVIRTLNIRVYVYALNKAESNLPSWDEWOTLAENMEAGLSTCOAOITLADYAE 940  
Db 906 QOLNVAPOGVSALVDLYIQSMKET-----PTTAQWENAGVLTAGLNSQOANTLHAFIDE 961  
Qy 941 RLSSVLCNFWLANIQPEGVLSHSDLDYSYFLIDNQVSSAIKTRRLABAIAGIQLYINRA 1000  
Db 962 SRSAALSTYIYIQAQAAKAAIKSRDDLQYLLIDNQVSAIKTRIAEALIASIQLYVNA 1021  
Qy 1001 LMRIBENARAADVSTQPFDTW--TVNNRYSTWGVSRVLYPPENYIDPTQRTGQTRMBEL 1059  
Db 1022 LENVENANSVSIQRFQFDMQKRYSTWAGVSQLVYPENYIDPTWIRIGQTRMDAL 1081  
Qy 1060 LENISOSKLSRDTVDAPKTYLTRPETVADLVKVSAYHDNVNSNGLTFWVGQTRLENLE 1119  
Db 1082 LQSVSQSLNADTVEDAFMSYLTSEQVANLKVISAYHDNINNDQGLTYFGLSETDAGE 1141  
Qy 1120 YVNRVNDISRMQAGLAAANAKWETKIDTAVNPPYKDAIRPVIFRERHLIIVKEEVAK- 1178  
Db 1142 YVRSVDHSGKFNDFKFAANAWSEWHKIDCPINPYKSTIRPVYKSLRYLLMLBQKEITKQ 1201  
Qy 1179 --NGTD--PVETYDFTKLAPLRHDGWSAPWSYDITTOVEAVYTDKPKDPTERLALAASG 1234  
Db 1202 TGNKSGQYQETDYRYELKLAHRYDGTWNTPTDVPNKKISEL--KLEKNRAPGLYACAG 1259  
Qy 1235 FQGEDTLVYVYKTKGYSDFGSKNKNVAGMTIYGDGSKFKNMENTALSRYSQLNTPDII 1294  
Db 1260 YQGEDTLVYMFYNOQDITLDSY--KNASMOGLYIFADMASKONTPEQSNVYRD--NSYQOF 1315  
Qy 1295 HTQGNDLVRKASYRPAQDPEVPASLNMGS--AIGDSDLTVMENGNIPQITTSYSSDNLAI 1352  
Db 1316 DTNN--VRRVNNRYAEDEIPESSVSRKDYCGWGYLLSMVYNGDIPITINYKAASDDLKI 1372  
Qy 1353 -----TLHNAFTVRVYDGSNGVTRNKQISAMKUTGVDDGKSQYGNAPFIANTVRYGGY 1405  
Db 1373 YISPKLRIHNG-----YEGQ----KRNQCNLNMYG-----KLGDKEIV-----Y 1409  
Qy 1406 SDLG--GPITVYVYKTKNY--IASVOGHLMNADYTRRLILTPVENNYARLPEF--PSPNTIL 1462  
Db 1410 TSLGVNPNNSNKLMPYVYQYSGNTSGLNQGR--LLFHRDTPYPSKVEAMIPGAKRSUT 1467  
Qy 1463 NTVFTVGS-----NKTSDFKKCSYAVDGNNSQGFQIFSSYQSGGLWIDTGINNTDIK 1515  
Db 1468 NQNAAGIDDYATDSLKNKPDLLKQYLFMTD--SKG-----TATDVSGPVEINTAISPQVQ 1520  
Qy 1516 ITWAGSKTHPTTASDHIALSPANSFDPAMPYTFKPLEIDASSLAFTNNTIAPLDIYFETKA 1575  
Db 1521 IIVKAGGKEQTPADKQVSIQSPSPFDEMNYQFNALEIDGSLNFNNNSASIDVTFAPA 1580  
Qy 1576 KQGRVLG-----KIKOTLSVKRVNYPEDILFLRETHSGAQYQMLGVYRIRLNTLASQIV 1631  
Db 1581 EDGRKLGYESFIPVTLKY-----STDNALTLHNNENGAQYQWQOSYRTRLNTLFAQLV 1635

Qy 1632 SRANTGIDTILMETQRLPEPPLGEGFPANFVLPKYDPAEHGDERWFKIHNIGVGNTR 1691  
Db 1636 ARATGIDTILSMETQIOEPQLGKFVATFVIPPYNLSSTHGDERWFKLYIKHVVDNNS- 1694  
Qy 1692 QPYYSGLMSDPTSEYSMTLF-----VPYAEGYMHGVRGVLGVGYQKITYDNTWESAFYFD 1746  
Db 1695 HIIYSQJTDN--INITLFIPLDDVPLNODYH-----AKYMTFKKSPSDGTWGWPHFVRD 1749  
Qy 1747 BTKQOVLINDADHSDGMTQOQIVKNIKKYKGLNLSIATGYSPAMDNSASALYWFEL 1806  
Db 1750 D--KGIIVIN-----PKSILTHFESVNLNIS-----SEPMDFSGANSLYWFEL 1793  
Qy 1807 YITPMFCQRLLOEKQFDEATOWINVYNPAGYVINGEIAPIWNCRPLEETTSNANPL 1866  
Db 1794 YITPMVLAQRLLEHQFNDEANRWLYKVMSPSGYIVHGQIQNYQMVNRPLLEDTSWNSDPL 1853  
Qy 1867 DAIDPDAVANDPWHYKIATFMBLLDQLILRGDMAYRELTRDALNEAKWYVYRTLELLGD 1926  
Db 1854 DSVDPDAVAQHDPMHYKYSTFMTLTDLLIARGDHAYRQLERDTLNEAKWYVYRTLELLGD 1913  
Qy 1927 EPEDYGSQOAAAPSLSGSAASQTVQAAQQDLMTLGRG---GVSKNLTANSVLGLFLPEY 1983  
Db 1914 KPYLPSTTWSDPLDRADITTONAHDSAIVALRQNIPTPAPLSRSANTLTDLFLPOI 1973  
Qy 1984 NPALTDYQTLRLRLFNLRHNLSDIGQPLSLAIYAEPTDPKALLTSMVOASQGSVAULPG 2043  
Db 1974 NEVMVNYQTLAQRYVNLRHNLSDIGQPLYPYIATPADPKALLSAAVATSQGGKLPES 2033  
Qy 2044 TLSLYRFPVMLERTNLVAQLTOFGTSLLSMAEHDDADELTLLLOQGMELATQISIRQQ 2103  
Db 2034 FMSLWRFPHMLNARGMVSQLTOFGSTLONTIERQDAEALNALLQNAEALLTLNLSIQD 2093  
Qy 2104 RTVDEVDADIAVLAESRRSAQNRLEKYQOQLYDEDINHGEQRAMSLDAAAGOSLAGQVLS 2163  
Db 2094 KTIIELEDAKTVLEKSKAGASRPDSYGLKYDENINAGENQAMTLASAAGLTTAVQASR 2153  
Qy 2164 IAGVADILVNVVFLGACGSGRWGAALRASAVMSLSATASOVSADKISRSEAYRRRRQEW 2223  
Db 2154 LAGAAADLVNPIFGAGGSRWGAATAEATGYVMEFSANVMNTADKISRSEAYRRRRQEW 2213  
Qy 2224 EIQDNADGEVQMDAQLESLEKIRREAAQOVVEYQETOQAOHTQAOLELLORKEFTNKALYS 2283  
Db 2214 EIQRNNAEALQIDAQKSLAVRREAAVLQKTSUKTOOQTSQSLAQFLQKFSNOALYN 2273  
Qy 2284 WMRGKLSAIYYQFPDLTOSFCMLMAQEARRELTONGVTFIRGGAWNGTTAGLMAGETLIL 2343  
Db 2274 WLRGELAIIYQFYDLAVARCLMAEQAYRWELNDDSAFIPKPGAWQGYVAGLLAGETLML 2333  
Qy 2344 NLAEMEKVWLBREDALEVTVTSLAQFYQALSSDN--FNLTEKLTQFLREKGNVNGASG 2401  
Db 2334 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKONGPFSLAQEDIKLVSOQSGSAGSN 2393  
Qy 2402 NELKL-----SNRQIEASVRLSDKI FSDYPESLGNTRPQLKOVSVTLPALVGPYSDIRAV 2456  
Db 2394 NNLAFAGCTDTKTSLQASVSFADLKIRIDYPASLGKIRRIKQISVTLFALLGPGYQDQAI 2453  
Qy 2457 LNYGSIWMPGCSAIALSHGVNDSQGMFLDNFSRYPFEGISVNDSGSLTLPSPDAT- 2515  
Db 2454 LSYGDKAGLANGCEALAVSHGWNDSQGFQLOFNDGKFLPFEGIAI--DOGTLTLSFPNASM 2512  
Qy 2516 ---DRQKALLBSLSDIILHIRYTR 2537  
Db 2513 PEKGQATMLKTLNDIILHIRYTIK 2537

Search completed: February 16, 2006, 21:31:20  
Job time : 263.656 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 56.6609 Seconds  
(without alignments)  
4309.818 Million cell updates/sec

Title: US-10-754-115-34

Perfect score: 13043

Sequence: 1 MYSTAVLLNKISPTRDQTM.....KALLESLSDIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_80.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	317	2.4	833	2 AB0448	probable insectic
2	254.5	2.0	4688	2 F92885	hypothetical prote
3	240.5	1.8	2893	2 A4556	toxin-like outer m
4	235	1.8	2178	2 S55805	alpha-toxin - Clos
5	229.5	1.8	2364	2 I40884	cytotoxin L - Clos
6	229.5	1.8	2383	2 D64962	probable membrane
7	222	1.7	2710	2 A37052	toxin A - Clostrid
8	222	1.7	5291	2 F90696	hypothetical prote
9	221.5	1.7	2902	2 C71953	toxin-like outer m
10	220	1.7	5005	2 F92884	hypothetical prote
11	220	1.7	5188	2 B85547	probable RTX famil
12	217	1.7	1830	2 E92909	conserved hypotet
13	216	1.7	2628	2 S59413	probable membrane
14	216	1.7	2529	2 B84635	toxin-like outer m
15	216	1.7	4385	2 T29042	hypothetical prote
16	211	1.6	2819	2 A90551	conserved hypotet
17	211	1.6	6713	2 B99921	hypothetical prote
18	205	1.6	1856	2 C95008	immunoglobulin A1
19	203.5	1.6	2491	2 A57036	talin - elime mold
20	202.5	1.6	3194	2 D71917	toxin-like outer m
21	201.5	1.5	2367	2 S70172	toxin B - Clostrid
22	201	1.5	1819	2 D97033	uncharacterized pr
23	200.5	1.5	1795	2 F97713	190K antigen precu
24	200	1.5	1628	2 E90538	hypothetical prote
25	199	1.5	2660	2 E85822	probable invasin Z
26	196	1.5	4199	2 S76412	hypothetical prote
27	195	1.5	3705	2 A00123	probable autotrans
28	194	1.5	1962	2 A32634	lactocepin (SC 3.4
29	194	1.5	3013	2 AB0480	probable invasin Y

#### ALIGNMENTS

##### RESULT 1

AB0448

Probable insecticidal toxin YPO3681 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C>Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C:Accession: AB0448

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0448

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-833 <R>

A:Cross-references: UNIPROT:Q8ZAV3; UNIPARC:UPI000000CDA68; GB:AL590842; PIDN:CAC93150.1;

C:Genetics:

A:Gene: YPO3681

Query Match 2.4%; Score 317; DB 2; Length 833;

Best Local Similarity 18.8%; Pred. No. 4.9e-09;

Matches 179; Conservative 151; Mismatches 314; Indels 310; Gaps 31;

Qy	49	LYHETIEQKNRLLERIFTRANPOLSGAIRLGIE-----RDSVRSYDEMFGARS	100
Db	19	IHQAKAEERHQAVRATQLRNDPVLGINKLAFSVAPKILQPEARDSLAEGIPERA	78
Qy	101	SSFYKPGSVASMFSPAGYLTLYREAKDLHFSSAYHLDNRRLADLTLSQSNMDTIS	160
Db	79	NEYADPASIQSLFSPGRYLCELYHVAKELHEDGNKLHDKRRPDLQELVLSNNMQVS	138
Qy	161	TLTUSNLLLEHITRTKGGSDALMESLS--TYRQADTPYHQPYETIRQVIMTHDSTLS	218
Db	139	SL-----BILLNVL-----QTNAPLAKLAKDTEAHANDVSFTLPY-----DDNLT	178
Qy	219	ALSRNPVMQAGASLAILANISPELYNLTITTEIKNADALFAON---FSEMITPEN	275
Db	179	VIN-----AILED-----KAISURETALLAENNDPWANPITP--	211
Qy	276	FASQSWIAKYGLBELSEVQKYLGMQLONGYSDSTSAYVDNIISTGLVNNESKLEAYKITRV	335
Db	212	-----ALVQEQGLNPASYA-----LIDIKSPL-----	234
Qy	336	KTDYDKNNINFDLMYEGNNOFFIRANPKVSREFGATLRKNAGPSGIVGSLSGPLIANTN	395
Db	235	--DD-----	236
Qy	396	FKSNVLSNISDSEYKNGVKIYAYRTSSTSATNOGGGIFTFESYPLTIFALKLKAIRLC	455
Db	237	-----NSAKRLA	243

Qy	456	LTSLGSPNELQTVIRSDNAQGIINDSVLTQVFTYFLYSH-----RYALSDDDAQVLNGSV	510
Db	244	HATOLSEVLOQLWLNKAIAESSDDKSLRPEILITIISEYRRLHQRYGLSVDPPIAINAV	303
Qy	511	INOYADDSVSHFNRLFNTPLPGKIPKADGNTVSIIDPEEQSTFARSALMRGLGVNSGR	570
Db	304	NTHYNENKTSFFQOIFST-----LDVDAGNFLD-----QGSW-EVIIRKALGITABE	351
Qy	571	LYQLGKLAGVLDQAQNTITLSVFVYSSLYRUTLLARVHQLTVNBLCLMYGL-SPFNGKTTA	629
Db	352	LLRIAKYCFGKSISNVKMSKKPSQLYRMAMIPRTILGVSFSAEYLMQLYSHPDENIME	411
Qy	630	SLSSGELPRL--VIMLYQVOTMLTEAEITTEAILWLLCTPEPSGNTSPETSINLNN---	682
Db	412	KIAGNALTIIIDAILVENTLQMSBQKIDITLQAMLTQYTTATPELNFSLNIYQT	471
Qy	683	-----LRPSIDEMASHNRELQAEILAPPIAATLHLASPDMARYILLMTDNLPRG	733
Db	472	LKGQVYESLKNLYRSLANGFH--LKANVAVGLVN--WLAKNDSEFTTLERFWONIS--	524
Qy	734	GLDIAGFWTLVLKESLNANETTO--LVQFCHVMAQLSLSVQTLRLSEAEISLVVISGEAV	791
Db	525	-----MTABEPSLHQLIHQPLLQ-COKLSQYVLIQAWELSQEIALILLPN---	573
Qy	792	LGAKNPAGQHNIDTFLSYRPHQWINGLGNPGSDTLDMLRQQTLT-----AD	839
Db	574	-GIDNRGSAPSPTILKLLSEFKLCOEAKVQSSELFDIMQOLIITDTNEKQEKLRNSAD	632
Qy	840	RLASVMGLDISWYTOAMVSAGVNLQCWQDINTVLQWIDVASAL-----HTMPSVIRTLVN	895
Db	633	KVIRSIKSIKGSINNSM-----DDISTISIRNGSATLFPPEHPM-----	672
Qy	896	IRYVTALNKAESNLPSWDEWQTLAENMEAGLSQQOATLADYTAERLSSVLCNW	949
Db	673	---YKALKLEVSNLEK-SKIQLEGKKKEEIKLEQAK-----DNIQSLINNW	715

RESULT 2  
 F82885  
 hypothetical protein UUA482 [imported] - Ureaplasma urealyticum  
 C:Species: Ureaplasma urealyticum  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: F82885  
 R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Caswell, G.H.  
 submitted to GenBank, February 2000  
 A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mir  
 A:Reference number: A82870  
 A:Accession: F82885  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-4688 <GLA>  
 A:Cross-references: UNIPARC:UPI0000110213; GB:AE002145; GB:AF222894; NID:g6899476; PIDN:  
 C:Genetics:  
 A:Experimental source: serovar 3; biovar 1  
 A:Gene: UU482  
 A:Genetic code: SGC3

[illegible]



Db 2302 TFFTEIINDNNKTFPFNKSDSIQOKFIVLNNQVGVGNIIEIQDROVNLHNSAKIRFEL 2361  
 Qy 1060 --LENIQSGLSRDVEAFKTYLTRPETADLVKVS---AYHDNVNSNTGLTWFGQT 1113  
 Db 2362 NDLNV---LSDN--EQATTYNNQTTSA--KVITDQOKYLEAFNSVLVN-----2407  
 Qy 1114 RENLPEYWRNVDIRMQAGELAANAKWKTKIDTAVNPYKDAIRPVIFRE-----RLHL 1168  
 Db 2408 -----KDTIINKIEFNTKPKNASKNIGINDT--NVIYDATNLIINDLKITGLPHT 2456  
 Qy 1169 IWVEKEVAKGTD-----PVETVDRFTLKLAPL-----RHDS--WSAPWYD--ITTOVE 1216  
 Db 2457 L---KEFEANNKNTISVSLELDTNNHISKNLYPIAKFSDNDQSVLTPNPIAADKIVTNNN 2513  
 Qy 1217 AVTDKPPTEBLAALASG---FOGEDTLLVVFVYTKGYSYDFGG---SNKNV-----1262  
 Db 2514 --NNKKELTFTNLWLTNSRQVIFKG-----LYVNSSNGTNDENKKEKNSVDYKITYK 2567  
 Qy 1263 -AGMTIYDGSF-----KMENTALSRYSQLKNTFDIHH-----TQGNLVR 1303  
 Db 2568 PTITTIQKNGWTFPPQNAQKFKFNINSDENVDFSTDLDAITTFSDQHDTKTITTKLKQ 2627  
 Qy 1304 KASRPAQDFEVPASLWMSAIGDSSLTMENG-----NIPQI---TSKYSSDLA 1351  
 Db 2628 KDNQWYIEDTISLAYNDTYKLDISITKPMNFAFANLKIQIENKBOISFTTQSGPTQLVS 2687  
 Qy 1352 ITLHNAFTVRYDGSNGVIRNKOISAMKLTGVDGKSOYGNAFIANTVKHYGGYSDL---1408  
 Db 2688 ITSDHLMNDQFNAS-----NQQTITAKVGVN--DLYNNRKI--KLVEYRSNNQKSLV 2738  
 Qy 1409 -GGPITVYNTKNTIASVOGHLMNADYTRRLILTPVENNYVARLFEPPFPSTNTILNTVFT 1467  
 Db 2739 ESNELTQKQTOYIFTLPISVANRQYSFKEIKTNNPFTL-----NNSTNVTSFN 2793  
 Qy 1468 VGSNKTDFKCSYAVDGNNGSQFOIFSSYOSSGWLDIDTGNNVDIKITWAGSKHTFP 1527  
 Db 2794 VNASKT-----QIVIDNTN-----LATNITSATITTYKLAKSKOHVF 2830  
 Qy 1528 TASHIASLPANSPDAMPYTPKPLEIDASSIAPNTNIAPLD-----IVPET---KAK 1576  
 Db 2831 QVGSIIITLKSND-----ERQESITYTKTITSISDGSSEATVSNFTDQTLKEE 2879  
 Qy 1577 DGRVLGKI---KOTLSKRVNYNPEDILFLRETHSGAQYMQL-----1615  
 Db 2880 ANYKLIKGVFKPTLAVTNINDANNVIF--EDNNSVNYNFKTLVDHKVTNVSSNDSTINT 2938  
 Qy 1616 -----GVYRIRLNTLASQLVSRANTGIDTILMTQRLPBPPLGEGFPANFVLP 1665  
 Db 2939 TTQTVNIDIDGIQRTWINKKI--QIVYTSNDGEEILSDQKTL-----LWAN---2982  
 Qy 1666 KYDPAEHGDEWPKHIGNVGNTGRQPYSGMLSDTSETMTLFPVPAVGYYMHGVR- 1724  
 Db 2983 -----NHYSFELNKHNRKYLUKEVRIINDNNKTSII-----FHLNGIAD 3024  
 Qy 1725 --LGVGQKITVDNTWE---SAFFYFDETKQOFVLINDADHDSGMTOQGIKVNKKYKGF 1779  
 Db 3025 WFIUNKTQSISSIIIEPIAKKNLQSTQIRFIL--NDPDVLSNEBEAL-----3073  
 Qy 1780 LNVSIATGYPAMPDNPNSASALYWEHFYFTYPMPCFQRLLOKQKQFDEATQWNYVNPAGY 1839  
 Db 3074 --INVGENLSVKAKVKGVSQKYLE-----TTFNNLALNQ--DTIINSITFINPKPIKA 3122  
 Qy 1840 IVNGEIAPIWNCRLPBEETTSNANPL-----DAIDP-----1871  
 Db 3123 ATNIGID-----NSNKIYNNNTN--NINPLKIDNNFSVIGVQSDTKRITLSEYNNKTNVKVN 3177  
 Qy 1872 -----DAVAQNDPMHYKIATFMRLLDQILIRGDMAYRELRDNLALNAKMYVRLLELLG 1925  
 Db 3178 LELKANDNILQN---LYPIAFKQONNETVAKSIVSYSSIKKQKQ-----IIEI--3225  
 Qy 1926 DEPEDYGSQWAAAPSLSGAASQTVQAAAY-----QODLTWMLRGGVGSKNLRNTANSILVGLFLP 1981

Db 3226 -----RIPNLKSNRLYHFDNLYYLLNQDDKDLIDNNKLIIRSNNVNVEI-----3268  
 Qy 1982 EYNPALTDY-----WOTLRLEFLNLRHNLISIDQOPLSLAIYABPTDPKALLTSMVQA--2033  
 Db 3269 EVKEGNTTFKQNVEMQ-----INSNSVNFKLQVB--SDDEDILDNNVMADV 3313  
 Qy 2034 -----SQGSASVLPGLTSLYR-----FPVMLETRNLVAQLTQ 2066  
 Db 3314 IFSSKSNQNDTKFVNNVVKLKEGNIWFIEGIVNNLKPETEYELKSIVLTKPLNANSNLT 3373  
 Qy 2067 FGTSLLSMAEHD-----DADELTTLLQQQMELATQSIRIQ-----QRTVDEV 2109  
 Db 3374 INPTNFKTKPGNYGIINIESNOTLTNSQNVNTIDGIRSEWPNKEAKLYVINTNDEV 3433  
 Qy 2110 DADIATLAESRRSAQRLE-----KYQOQLYD-----2135  
 Db 3434 EAKVNL-----EENKLVYVFNFNKLYNRKYTKLSIKWFDQGNWIDFPNKNQNLQNSFI 3486  
 Qy 2136 -----EDINHEGORAMSLDAAAGQ--SL--AGQVLSIAEGVADLVPVFGACGGS 2183  
 Db 3487 IKGKTPISVNNKLANEISDRVINNLNVAKIQFSLNDPEDVLSNNE--IAEIT-----3534  
 Qy 2184 RWGAALRASASV-----MSLSATASOVSADK--ISRSEAYRRRRQSEW---IORDNAD 2231  
 Db 3535 -YDNKKTNAKVIITVNNKYLEVLSNLEFNFHDTIINKIEFITRPNKAVDSVGNNNSKI 3593  
 Qy 2232 GEVQMDAQLLESKIR-----REAAQWQVEYQETOQAHTQAO 2268  
 Db 3594 YELTNLIBETRKLIINDFRISGLSSNNLHLEIKYNDKESINSSTIEF--TINKHIAKN 3651  
 Qy 2269 LELLQRFPTN-----KALYSWNRGKLSAIYY-----O 2295  
 Db 3652 LK--FKLFRPEFGHEVYSELIDGNINNKTKQEVLFVSYGLKSNQLYKLVVYVYLDNTHQ 3710  
 Qy 2296 PFDLTQSFCLMAQALARELTDN--GVTFI--RGGAWNGTTAGLMAGETLLLLAEME--KVM 2352  
 Db 3711 NIDTRK--IPKHNVSKHIEINPGVTWISKHGWNKSPD-----TTANPEFKIE 3758  
 Qy 2353 LERDERAL--EVTRTVSLAQFYQALSSDNFNUTEKLTQFLREG-----KGNVAGSAGNELKL 2406  
 Db 3759 TQDDNDVLANIDAIVK-----FKDEHNNIKQIVRIKENNDWLKIQIDNLNPETKY 3811  
 Qy 2407 SNROIE-----ASVRLSD-----LKIFSDYPESLCINTRLQKQSVT 2442  
 Db 3812 KLENIELSKPLKTHNLSVINDKENISLITETGNPVLKVIQTQNDTINDTQO--TINTV 3869  
 Qy 2443 LPALVGVP--EDIRAVLNYGSIWMPRGCSAIALSHGVNDSGQFMDFDNDKRYLPFEGIS 2500  
 Db 3870 LSGVNSKNGRQIKVYVKDNNVIYE--SSLITLQKKNQYQLLSNLSNREYRFEKIE 3927  
 Qy 2501 VNDSGSLTSPDADTRQKALLESLSDI 2528  
 Db 3928 INHI--SNTNNPFD-----LEKLVGV 3946

RESULT 3

A64556  
 C:Species: Helicobacter pylori  
 C:Date: 09-Aug-1997 #sequence\_revision 09-Aug-1997 #text\_change 09-Jul-2004  
 C:Accession: A64556  
 R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 388, 539-547, 1997  
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.  
 A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.  
 A:Reference number: A64520; MUID:97394467; PMID:9252185  
 A:Accession: A64556  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-2893 <TOM>  
 A:Cross-references: UNIPROT:Q25063; UNIPARC:UPI000000D3097; GB:AE000547; GB:AE000511; NID

Query Match		1.8%; Score 240.5; DB 2; Length 2893;
Best Local Similarity		18.6%; Pred. No. 0.0067;
Matches 361; Conservative 275; Mismatches 745; Indels 563; Gaps 79;		
Qy	56	QKQNRRLERARIFRANPQLS-GAIRGIERDSVRSYDEMFGARSSFFVKPGSVASMF5 114
Db	1234	QAQNSYFIDDTAQQGNGFNFSNALNL---DPSNSFRGVYGVKTQSVFRNKAQNAISFT 1289
Qy	115	PAGVILTE-LYR-EAKDLHFSSAYHLNRRPDLDLTLQSNDMDTEISTILSNEILL-- 170
Db	1290	NSTNLSSGLYQMAQSFLFONS-----NLSVSVGTSSIKANAINLSQNASINA 1337
Qy	171	-EHTTRKTGSDAL-MESISTYRQADTPYHQPVTIROVIMTHDSTLS---ALSNP 225
Db	1338	SNHSTLEBQDLNVDTSLLMNOSTINVSNATINDYASLIASGHLPNGAVNFNS- 1396
Qy	226	VMGOAGASLAILANISPELYNLTBEITEKNADALFAQ-NFSEN-----ITP- 273
Db	1397	-----ANITTSLNN---SSIIVFGAVSLGGQFNLSNNSSLDFOGSSAITSN 1439
Qy	274	-----ENFASQSWIAKYGLESEVOKYLGMLQNGYSDTSATVNMISTGLVNNESKL 327
Db	1440	TAENFYDNAFSQSPITTHQALDIKAPLSLGGNLLN---PNNSSVLDLKNQSLVFGDQSL 1496
Qy	328	EAYKIVTRKTDYDKNINYPDLMEVGNQPPIRANEKVSREFGATLR-----KNAGPSG 381
Db	1497	NIANIDLSDLNKNNRVYNIQADMNSWYERISF-----FGHIINDGIYDAKNQYYS- 1550
Qy	382	IVGSLSGPLIANTNFKSNYLS-----NISDSFYKNGKVIYAYRTSSTSATNQGGIF 434
Db	1551	FTNPLNALKITESFKDQNLQSVLTSQIPGIKNTLYNIGSEIFNYQ-----KVYNNANGVY 1605
Qy	435	TFESYPIITIPALKUNKAIRCLTSGSPNELQTVIRSDNAGGIINDSVILKVFYTLFYSH 494
Db	1606	SY-----SDDAQGVPLYTSNKGYY----- 1625
Qy	495	RYALSPDDAQVLINGSVINOVADDSVSHENRLENTPLPKGIPEADQNTVSIIDPDEQST 554
Db	1626	-----NPNQSVQASGNNNTKNNLTSESSIIISQTYNAQNPIS----- 1664
Qy	555	FARSALMRGLVNGSELYQGLKAGVLDQAQ-NTITLSVFVISSLYRLTLARHQLT--- 610
Db	1665	---ALHIYKNGYNFNKIKALQMAKLYPEIKKVLGNDFFSPLNALNSNA-LNQLTKLI 1720
Qy	611	-----VNELCMLYGLSPFNKKTASLSGSELPRVLWLVQVQWLTAEAITTEALWLL 663
Db	1721	TPNDWKINEL-----IDNANNVYQNFNG---TLIVGATQIGQTDNTSAVVFGLGYQ 1772
Qy	664	CTPEFGNISPEI-SNLLNLRPSISSEDMAQSHNREIQAETLAPFAIATLHLASP----- 717
Db	1773	TPCDYDILVQKFRGTGLQLLESSADLGVIDTTFNAKEI---YLTGTLGSGNANGTGG 1829
Qy	718	-----DMARYILLWTNLRPGGLDIAGFWTLVLKESLN-----ANETQL-VQFCH 762
Db	1830	SASVTPNSQTSLLILNQANIVSSQTD--GIFSLMQEGINKVFNQAGLANILGEVAVQSI 1887
Qy	763	VMAQL-SLSVOTLRLSBAELSVLVISGFAVLGAKNP-----AGQHNIDTFLSLYRFHQWI 817
Db	1888	KAGGLGNLIIVNTLG-----SNSVIGVLYTPEQKNQTLSQLQGNFNDL----- 1931
Qy	818	NGLNPGSDTL--DMLRQQ-----TLTADRLASVGLDLSMVTQAMVSAGVNOQCWODI- 870
Db	1932	--MNDSGLNTAIKDLIKQKGLFWTGLVGLGAGLGIGDILQNPKEKIGMSINDLSKXGLF 1989
Qy	871	NTVLQWIDVASALHTMPSVIRTLNIRYVVTALNKAESNLPSWDEWQTLAENMEAGLSTQQ 930
Db	1990	NQITGFISANDIGQVISWMLQDIK-----PS-----NALKNDVAALGKQM 2030
Qy	931	-AQTADYTAERLSSVLCNFWLANIQEGVSLHSDLDLYSYFLIDNQVSSAIKTRTLAEA 989
Db	2031	IGFPLQDQTLNSLESLLQNOQIKSVLDKVAAGKLGPIYEQGLGD-----LIPN 2079

Qy	990	IAGIQLYINRALNRI-----EPNARADVSTROFFDTMTVNNRYSTWGGVSRVLYVYENY 1043
Db	2080	LGKKGLFAPYGLSQVMQKGFDSFNAQGNFV-QNSTFSNANGGTLSSNAGNSLFIAGNNH 2138
Qy	1044	IDPTQIRIGTRMDELLENISQSKLSRDYVEDAFKYLITFEVTADILKVVSAHYHDVNSN 1103
Db	2139	IAFTNHAGTLQLLSQDVSNIN-----ITTLNASNGLUK-INAAANNVSVS 2181
Qy	1104	TGLTWVF-----GQTRNLPYVWRNVDIRMOA- 1132
Db	2182	QG-NLFWASCAQSDPTTANIANPCALSAQSTNGASSNANNPATALSNNDESLMVA 2240
Qy	1133	-----GELAANAKWKWKIDTAVNPYKDAIRPIFRERLHLINVEKEEVAKNGTDPVET 1186
Db	2241	NDFNFSGNIYANGVWDFSKIKGSAN----- 2265
Qy	1187	YDRPTLKLAFLRHDDGSWSPWSYDITTOVEAVTDKPD--TERLALAASQFQGBDITLLVF 1244
Db	2266	-----IKNLYLYNNAQFOA---NNLTISNOAVLEKFNASFTVNNLNIQGA-FNNNATQKIE 2316
Qy	1245	VYK-----TGKSYSDFGSSKNVAGMTIYDGSFKKMENTAL-SRYSOLKNTFD 1292
Db	2317	VLQNLVIASNASTGIYGLEVGALNNSGAIHP-----NLNENTQTPPLIOAQEIIIN 2369
Qy	1293	IIHTQ-----GNDLVRKASYRFAQDFE-VPASINMGSAIGDDSLTYMENGNIPOITSK- 1344
Db	2370	LNTTQTPFMVNNSMANNITTYLLKSSRYIDYINPNLSQSYLNLYTLININGNHIIEKN 2429
Qy	1345	YSSDNIAITLHNAFTVRYDGGNVIRNQKIS-----AMKLT----- 1381
Db	2430	GALTYLQORVLLQDKLLSVALPNSNNSQNNILSLSVLYNQVMSCGDKAMDFTPTTL 2489
Qy	1382	-----GVDGKSQY-----GNFIANTIKVHYGSDLGPIITVYKNTKNIYASVQCHL 1429
Db	2490	QDIYVIGQGSALNQIEAVGCGNAIKMLSTLMETKENPPFAPYILKXHSNEILGVTKDL 2549
Qy	1430	MN-----ADYTRRLILTPVNNYIYARLFEFFPFPSPNTIILNTVFTVG 1469
Db	2550	QNTASLISNPNFRDNATNLELASYTQOTSRITKLSDFRSREGESDFSLELKNKRFSDP 2609
Qy	1470	SN-----KTSDFKK-----CSYAVDGNM-----SQGFQ-----IPSSYQSSGW 1502
Db	2610	NPEVFKVYSQLSKHPNMLWQGVGGASFISGNGTLYGLNAGYDRLVKVNLGILGYVAYG 2669
Qy	1503	LDIDTGI-----NNTDILKITVMAGSKTHFTTASDHIAASLPANSFDPMPYTFKPLEIDASS 1557
Db	2670	SDFNGNIMHSLGNVDVGMVAPLAKNEFT-----LSAN----- 2704
Qy	1558	LAFTNNIAPLDIVFETKAKDGRVLGKIKQTLV--KRVYNP-----EDILFLRE 1605
Db	2705	-----ETVGNATSINSNLSVLNQRYNTWTTSVNGVGYDFMFKQK 2750
Qy	1606	T-----HSGAQYQMLGVYRIRLNTLLASQLVSRANTCIDILTMETORLPEPPLGEG--F 1658
Db	2751	SVVLKPOGLSYHFHIGLSGKMGKNDAAKQFLMHSNPNSESVTLNMLGSRKRYFGKNSYY 2810
Qy	1659	FANFVLPKYDPAEHGDERWFKIHGNGVGTGQPPYSGMLSDTSETSMTLFVPEYAGYY 1718
Db	2811	FVTARLGR-----DLIKSKGSNTVRPVGENTLILYRKGEVNT--FASVITGGE 2857
Qy	1719	MH-----EGVRLGVGYQKI 1732
Db	2858	MHLWRLVYVYVAGVGLKXGLQYQDI 2881

RESULT 4  
 S55805  
 alpha-toxin - Clostridium novyi (ATCC 19402)  
 C/Species: Clostridium novyi  
 A/Variety: ATCC 19402  
 C/Date: 28-Oct-1996 #sequence\_revision 08-Nov-1996 #text\_change 09-Jul-2004  
 C/Accession: S55805; S71294; S71158; S4273; I40834; S44272  
 R/Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.



Db 1666 KDLNLYIDSVIIIPKVVTKIIHPIPNPQINIINTQSIDHKCHLI-----IDSVLTN 1720  
Qy 1540 SFDAMPYTFKLEIDASSLAFTNNIAPLDIVFETKAKDGRVLGKIQTLSVKRVNYP-- 1597  
Db 1721 NYH-----WESDGDLLIITNG---LDI--NIRILQGLSFGPKYNIYKFSNYDELS 1767  
Qy 1598 -EDILFLRETHSGAQM-----QLGVYRIRLNTLLASOLVSRAANTGIDITIL 1642  
Db 1768 LNDLFQNYNVKGLIYINGELHYKNIIPGDTFEGWINIDSRWYFDSINLIKAGYQEI- 1826  
Qy 1643 TMTQRLPEPPLGEGFFAN-----FVLP---KYDPAEH-GDERWPKIHIGNYGGN 1688  
Db 1827 -----EGERYFNPNTGVQESGVFLTPNGLEYFTNKHASSKRW----- 1864  
Qy 1689 TGROPYSGMLS-----DTSETMTLFPVPAEG-YMHGEGVRLGVGYQ-----KIT 1734  
Db 1865 -GRAINTGWLTDGNKYFYFOSNSKAVTGLQKISDKYFYFNDGQMQIKWQIINNKKYF 1923  
Qy 1735 D-NTWESAFFYFDETKQFV-----LINDA-----DHDGQMTQGI 1769  
Db 1924 DQNTGEAIIHFNKKERYFDSRGLLTGVQVIGDKSYFSDNINNGWEGSGVLKSGI 1983  
Qy 1770 VKNIKKYGFLNV--SIATGVSAPMDFNSASALYYW-----LPYTPMM 1812  
Db 1984 FKTPSGFKLFSEGDKGAINYKGLWDLGNK--YFNSDSIAVTGSYNIKGIQYFNPKT 2041  
Qy 1813 CFQRLLEKQF--DEATQWYNYVNPAGY 1839  
Db 2042 A---VLTNGMYTLDDNNYVYVSGHNVLGY 2067

RESULT 5  
I40884  
cytotoxin L - Clostridium sordellii  
C;Species: Clostridium sordellii  
C;Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 09-Jul-2004  
C;Accession: I40884  
R;Green, G.A.; Schue, V.; Monteil, H.  
Gene 161, 1995  
A;Title: Cloning and characterization of the cytotoxin L-encoding gene of Clostridium sordellii  
A;Reference number: I40884; MUID:95369733; PMID:7642137  
A;Accession: I40884  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2364 <RES>  
A;Cross-references: UNIPROT:Q46342; UNIPARC:UPI00000B02BE; EMBL:X82638; NID:g1000694; PI00000694  
C;Keywords: cytotoxin

Query Match 1.8%; Score 229.5; DB 2; Length 2364;  
Beat Local Similarity 18.5%; Pred. No. 0.0018;  
Matches 429; Conservative 308; Mismatches 711; Indels 873; Gaps 119;

Qy 23 ADLOYLSFSELKTIFFDQLSGWRARHLYHETIEOKNNRLLERARIFTRANPQLSGAIRLG 82  
Db 7 AOLQKVVVFPRIQEDYVAINALALEYHNMSSESVVEKYLK----- 50  
Qy 83 IERSVRSYDEMGARSSSPKGVASMPSPAGILT-----ELYREA-----KDIHF-- 131  
Db 51 -DINNLTNDY-----LNTYKSGRNKALKKFKXYLTMEVLELKNNSLTPEVKNLHF 102  
Qy 132 -----SSAYHLNRRPDLADLTLSQSNMDPEISTLTSNELLLEHITRKTGSDSALME 186  
Db 103 IGGQINDTAINYNQKDV-----NSDYTVKVFVDSNAFLNTL-KKTIVES-ATNN 152  
Qy 187 SLSTYQAIDTP-----YHQPETIRQVI---MTH---DSTLSALSRNPVEMGQAGASLLAI 238  
Db 153 TLESFRENLDNPEFDYKFKRMEIIVDKQKHFIDYKSGIENPPEF-----I 201  
Qy 239 LAN-ISPELYNLTETI-----TEKNADALFAQNFSENITPENFASQMTAKYVLEL- 290  
Db 202 IDNIITKTVLSNEYSKDEALNKYIEESLUNKITANGNDIRNLEKFADED-LVRVYNQELV 260  
Qy 291 -----SEVQKYLGLMQLNGYSYSTSAYVD-NISTGL-----VVNNE----- 325

Db 261 ERWNLAAASDILRISMLKEDG-----GVYLDVDILPGIQDLPFKSINKPDSITWTSWEMI 315  
Qy 326 KLEAYKITRVKTDY-----DKNINYFDLMVEGNNQFFIRANFKVSRFEGATLRKAGPS 380  
Db 316 KLEAI-----MKYKEIYIPGYTSKN-----FMDLDE-----EVQSFESALSCKDKS 357  
Qy 381 GIVGSLG---GPL-----TANT-----NPKSNYLSNISDSEYKNGKVIYAYRYTSS 423  
Db 358 EIFLPLDDIKVSPLEVKIAFANNNSVINQALISLSDYCSLDVINQIKRKYKILNDNLNPS 417  
Qy 424 TSATNOGGGIFTFESYPLTIFALKL-----NKAIRLCILTS-----GLSPNELQTIYRS- 471  
Db 418 ---INEGTDFNT---TMKIFSDKLASI SNEDNMFMKITNYLVKGFAPDVRSTINLSG 470  
Qy 472 -----DNAQGI-----INDSVLTKVFPYTLFYSHRY----- 496  
Db 471 PGVYTGAYQDLLMPKDNSTNIHLEPELRNPFEPKTKISQLTQEITSLMSFNQARAKSQ 530  
Qy 497 -----ALSFFD---AQVLNGSVINQVADDSVSHFNRLFTNPLPKGK----- 535  
Db 531 FEEYKKGYPGALGEDNLDFAQ-----NTVLDKDYVS--KKILSMKTRKEYIHYI 581  
Qy 536 -----IFEADGNTVSDPDEE---QSTFARSALMRGLGVNSGELYQLGKLAGVLDAQ 584  
Db 582 VOLQGDKI SYEASCNLFSDPYSSILYQKNIEGSETAYVYVVADEAIKEIDKYRIPYQIS 641  
Qy 585 NNTLSVFISSLYRLTLARVQLTVNELCMLYGLSPFNGKTTASISSELPLRLVILWLY 644  
Db 642 NKRNI-----KLTFIG-----HGKSEFNTDTPANLDVDSLSSEIETIL 679  
Qy 645 QVTQWLTEARITTE--AIWLLCTPEPSGNISPE-----ISNLNLNLRPISDIMA 692  
Db 680 N-----LAKDISPKYIIEINLLGCMWFYSISARETYPKLLKIKDRVSELMPSISQDSI 735  
Qy 693 QSHNRELQABILAPPTAATLHLASPDWARYIL---LWTDNLRPGGLDIAGFMVLVLKE- 747  
Db 736 TVSANQVEVRI-----NEEGREILDHSGKWINKESIIKDISS-----KEY 777  
Qy 748 -SLNANETTQIV--QFCHVMAQLSVQTLR-----LSEAELSVL----- 784  
Db 778 ISFPNKENKIIVKSKYLH---ELSTLQEIERNNANSSDIDLEKKVMLTECEINVASNDR 834  
Qy 785 -VISGPAVLGAKQOPAGQHNI-----DTLPSLYRPHOWINGL----- 820  
Db 835 QIVEG-RIEEAKNLTSDSIYKNEFKLIESISDSLYDLK--HQ--NGLDSSHFI SFEDI 889  
Qy 821 -----GNP-----GSDTL-----DMLRQQT 835  
Db 890 SKTENGPRIRPINKETGNSIFIETEKEIFSEYATHISKEISNIKDTIFDNVNGKLVKVN 949  
Qy 836 LTA-----DRLASVMGLDISWVTOA---MVSAGVNOLOCMQDINT 872  
Db 950 LDKAAHEVNTINSAPFIIOSLIETNYTTKESLNSLVAMKVQVYVQALFSTGLNTI----TASK 1006  
Qy 873 VLOWIDVA---SALHTMPS-----VIRTLVN-IRYVLTALNK-ABSNI-PSMDEWOTLAENM 922  
Db 1007 VVELVSTALDETDLTLSEGLPIIATIIDGVSGLGAAIKELSETNP-----LLRQEI 1060  
Qy 923 EAGLSTOQAOTLADYTAERLSSV-LCNWFLANIOP-EGVSLHSRDDLYSYFLIDNQVSSA 980  
Db 1061 EAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKV 1120  
Qy 981 I---KTRTIAEATAGIOLYINR-----ALNRIEPNA-----RADVSTROFF 1018  
Db 1121 IDYFKHISLAEAGETFLDDKIIMPQDVLVLSIDFNNNSITLKGKCEIWRAGSGSHTL 1180  
Qy 1019 TDMTVNNRYST-----WGGVSRVLYYPENYID-----PTQIRIGOT----- 1053  
Db 1181 TD-DIDHFFSPSPITYRKPLSIYDVNLNIKKEIDFSKDLMLVLPNAPNRVFGYEWGTPG 1239  
Qy 1054 -RMD-----ELLENISOSKLSRDTVEDAFKTYLTRFETVAD---LKVVSAYHD-NVNSNT 1104

Db 1240 FRSLDNGTKLLDRI-----RDHYEQF--YMYFAFIADALITKLPYEDTNRINL 1291  
Qy 1105 GLTWFGVQTRNLPEYWRNVNIDSRMQAGELAAWKEWTIDTAVNPKDAIPFVIRE 1164  
Db 1292 D-----GNTR-----SFIQVPI--- 1303  
Qy 1165 RLHLIWEKEVAKNGTDPVYDRFTLKLAFRLHDSWSA-----PW 1207  
Db 1304 -----TTEQIRKN-----LSYFYSGSGSYSLSPYNNMIDNLVENDTW 1344  
Qy 1208 SYDITTOVEAVT---DKKPDTERLALASQGEDTLVFFVYKTKSYSDSGGSKNVAG 1264  
Db 1345 VIDVNVVKNITISDEIQKGLIENILSKNIENDKILNNHTFNFYGDINESNRFTS- 1403  
Qy 1265 MT-----IYDGSFKKMENTALSRYSQLKNTFDII----- 1294  
Db 1404 LTFSEILEDINIIIEIDLVSYSKILLSGNCMKLIENS-----SDIQKIDHGFNGEHQK 1458  
Qy 1295 -----HTQGNLNV--RKASVRFPAQDF-----VPASLNGMSAIGDD--SLTV 1332  
Db 1459 YIPYSIDNETKYNGFIDYSKKEGLFTAEPFSGNESIIRNIYMPDSNNLFIYSSKDKDIRI 1518  
Qy 1333 MENGNIPOITSKSSDNLAIITLHNAAFVRYDGGSGNVIRNKQISAMKLTGVGDKSQYQNA 1392  
Db 1519 INRGDVKLLIGNYFKDKDKVSL-----SFTI-----EDTNTIKLVG-----YLDE 1559  
Qy 1393 FIANTVKHGYGSDLAGGPITVYNTKKNYIASVOCHLNMADYTRRLILTPVENNYARLF 1452  
Db 1560 NGVAQILKF---MNAKSALNTSLSLNNFLSIN-----IKNIFYNNL- 1599  
Qy 1453 EFPSPNT--ILNTVTV--GNKTSDF-----KKCSYAVDGNNSQG 1490  
Db 1600 ----DPNIEFLDTNFIISSGNSIQFELICDKDKNIQPYEFNKIKETSITLYVGNRQN 1655  
Qy 1491 FOIFSSQSSWLDIDTGINNTDIKITVMAKSKHTFTASDHASLPANSDAMPYTKP 1550  
Db 1656 LIVEPSYH---LD-DSG---NISSTVINFSQKLYGIDRYV-----NKVIIAPNLYTD 1701  
Qy 1551 LEIDASSLAFTNTAPLDIVETRAKQGRVLGKIKQTLVSRVNVNVEDILFLRETHSGA 1610  
Db 1702 -EINITPYKNYICPEVILIDANYINEKI-----NVINDLSIRYVD----- 1744  
Qy 1611 QYMQGVYRIRLNTLLASQLVSRANTGIDTILTWETQRLPEPPLGEGEFANFVLPKYDPA 1670  
Db 1745 -----NDGSDLLILANSEEDNQPV----- 1764  
Qy 1671 EHGDERWPKIHGNGVGTGQPYSGMLSDTSETMTLFV-----PYABGYMYHE--- 1721  
Db 1765 -----KIRFVNFKSDTAADKLSFNFSQKQVSVSKIISTFSLAAYSDFDYBFL 1816  
Qy 1722 -----GVRLGVGY-----QKITDONTWESAPFYDETKQQFVL 1754  
Db 1817 VSLDNDFYNSFGNMGVSLIYINDSLYFKPKNNLITGETTIDGNKYFYDPTKSGNAS 1876  
Qy 1755 INDADHDGSG---MTQQQGVKNKYKGFNLVSIATGYSAPM-----DFN 1795  
Db 1877 IGEITIDGKYFNKQGILO-----VGVIINTSDGLKYPAPAGATLDENLEGESVNFIGKLN 1931  
Qy 1796 SASALYTWELPYTPMCMQFRLLOEKQFDEATQMINVYNP 1836  
Db 1932 IDGKIYYPEDNY-----RAAEVEMKLLDDET-----YFNP 1961

RESULT 6  
D64962  
probable membrane protein b1978 - Escherichia coli (strain K-12)  
C:Species: Escherichia coli  
C:Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: D64962  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: D64962  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2383 <BLAT>  
A:Cross-references: UNIPARC:UPI0000047C86; GB:A5000289; GB:U00096; NID:q1788285; PIDN:A  
C:Keywords: nucleotide binding; P-loop; transmembrane protein  
F:54-70/Domain: transmembrane #status predicted <TM>  
F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 1.8%; Score 229.5; DB 2; Length 2383;  
Best Local Similarity 18.1%; Pred. No. 0.0019;  
Matches 491; Conservative 352; Mismatches 880; Indels 983; Gaps 127;  
Qy 208 QVIMTHDSITLSALSRNEVVGQAGSLLATLANISPELYNLTETETKKNADALFAQNF 267  
Db 85 QIAIANNTV-----PYTLGALESAQSA-----ERFGISVAELKLNQFRTFARSF 131  
Qy 268 -----SENTIPENFASQSWIAKYKYLESEVQKVLQNGYSDSTSA- 310  
Db 132 DNVRGQDELDPVAPQVSEKLTTPPGNSSDNLEQ-----QIASTSQIGSLAEDMNSQAA 187  
Qy 311 -----YVDNISTGLVANNESKLEAYKTRVKTDDYDKNINVFELM---YEGNQOFFIRAN 362  
Db 188 NMARGWASSQASGAMTDLRFGTARITLGVDEDFSLKNSQDFLHPMYETPDNLFFSQH 247  
Qy 363 PKVSREFGATLRKNAGSPGIVGSLGGLIANTNFKSNISDSSEYKNGVKIYAYRYTS 422  
Db 248 -----TLHRTDERTQINNGLW-----RHPTPTWMSGIN-----PFDDHDL 283  
Qy 423 STSATNGGGIFTPESEYPLTIFALKKAIKRLCTSLGSPNELOTIVRSDNAQG--IIND 480  
Db 284 SRVHSRAGIAEYWRDY-----LKLSSNGYLRLTNWRSAPELNDYEAPANGWDVRAE 337  
Qy 481 SVLT-----KVPYTLFYSHRYALSFDDAQVLNGSVIN---QYADDDSVSHENLENTP 530  
Db 338 SWLPAWPHLGKLYEYGYDEVALFDKDRQSNPHAITAGLANTYTPPLMT-----FSAE 392  
Qy 531 PLKGIKFEAD-----GNTV--SIDPDEBQSTFARSALMRGLGVNSGELYQLGKL 577  
Db 393 QRQKQKQENDTRFAVDFTWQFSGAMQQLDNE---VAARSL-----AGSRVYL--- 439  
Qy 578 AGVLDAQNTITLSVFVYSSLYRLTLARVHQLTVNELCMLYGLSPFNKTKTASLSSGELP 637  
Db 440 ---VDNRNNIVLE-YRKKELVRLTL-----TDPVTGK-----SGSVK 472  
Qy 638 RLVIWL-----YQYQMLTEAE-----ITTEAIWLLCTPEFSGNISPEISNLLNLRP 685  
Db 473 SLVSSLOTKYALKGYNVEATALEAAGKVVTGKDILVTLPAYRFTSTPETDN----- 525  
Qy 686 SISEDMAQSHNRELAQAEILAPFIAATLHLASPDMAVYILLWTDNLRPGGLDIAGFMTLVL 745  
Db 526 -----TWPIEVTAE-----DVKG----- 538  
Qy 746 KESLNANETTLQVQFCHVMAQ-----LSLSVQTLRLSEALSVLVISGFAVLGAKNQPAQ 801  
Db 539 --NLSNREQSMVVQVAPTLTSQKSSVLSITOTLN-----ADS 573  
Qy 802 HNIDTFLSLRFHGWINGLGNPGSDTLDMLRQQLTLADRLASVMGLDLSMTVQAMVSAGV 861  
Db 574 HSTATLTFFIAH-----DAAGNP-----VVGLVLSTRHE-----GV 603  
Qy 862 NQ--LOQWQD-----INTVLQWIDVASALHTWPSVI-----R 891  
Db 604 QDITLSDWKNGDGSYTYQILTGTGAMSGTLTLPQLNGVDAKA---PAVVNILLSVSSR 659  
Qy 892 TLVNI-----RY-----VTALNKAESNLPSPDWEQTL-----AENMEAGLSQQAOTL 934  
Db 660 THSSIKIDKRYLSGNPIEVTVELRENDKPKVQKQQLNNAVSIDNVKPGVTIDWKETA 719  
Qy 935 ----ADYTAERLSS-----VLCNWFLANTQPEGVSLHSRDDLYSY-FLTDNQVSSAIK 982



[illegible]









I: Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;  
 Nature 397, 176-180, 1999  
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path  
 A:Reference number: A71800; MUID:99120557; PMID:9923682  
 A:Accession: C71953  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2902 <ARN>  
 A:Cross-references: UNIPROT:Q9ZME6; UNIPARC:UPI0000035923; GB:AE001464; GB:AE001439; NID  
 C:Genetics:  
 A:Gene: jhp0274

Query Match 1.7%; Score 221.5; DB 2; Length 2902;  
 Best Local Similarity 18.1%; Pred. No. 0.0072;  
 Matches 353; Conservative 301; Mismatches 718; Indels 583; Gaps 85;  
 56 QKQNNRLIARIFTRANPOLG-GAIRLGIERSVRSYDEMFGARSSSFVKPGSVASWFS 114  
 1241 QAQNSYFIDDDTAQNGNPFNSFALNL-----DFNSSEFVGIVGQTSVFKFNVAIVSFT 1296  
 115 PAGYLTE-LYR-EAKDLHFFSSAYHLNRRPDLADLTLSQNMDEISTLTLSNELLL-- 170  
 1297 NSSNLSGLYQWQAKSVLFDNS-----NLSVSVGTSSSIKANAINLSQNASINA 1344  
 171 -EHITRKTGGD---SDALMBESLSTYRQALDTPYHPYETIRQVITWHDSTLS-----ALSRN 223  
 1345 SNHSTLEQGLNLNDT---SSNLNQSAINVSNNATINDYASLTASNGSHLNFNGAVNFN 1402  
 224 PEVMQAEASLLAILANISPELYNLTETEKNADALFAQ-NFSEN-----IT 272  
 1403 S-----ANITTSL-----SSSIVFAGVSLRGQFNLNNSSLDFOGSSAIT 1444  
 273 P-----ENFASQSWIAKYGLSEVQKYLQMLQNGYSSTGAYVDNISTGLVNNES 325  
 1445 SNTAFNFYDNASQSPITFHQALDIKVLPSLGGNLLN---PNNSVSLNKLNSQLVFSDOG 1501  
 326 KLEAYKTRVKTDDYKNNINFDLMYEGNNOFFTRANKVSRERFATLRKAG----- 378  
 1502 SLNTANIDLLSGLNKRVRVNIIOADMNGWYERINF-----FG--WRINDGIYDAKQ 1554  
 379 PSIGVLSGLPIANTNFKSNYLS-----NISDEYKNGVKIYAYRYTSTSTSATNQG 431  
 1555 TYSFTNPLNALKITESPKNNQLSVTLISQIPGKNTLYNISSEIFNQ-----KVYNNAN 1609  
 432 GIFTPEYPLTIFALKUNKAIRLCLTSGLSNPELOTVRSNAGIINDSVLTKVFTLP 491  
 1610 GVYSY-----SDAQGVFLTSSVKGYY--- 1632  
 492 YSHRYALSFDDAQVLNGSVINQYADDDSVSHFNRLFNTPLPKGIFFADGNTVSDIDPEE 551  
 1633 -----NPNQSQAGSNNTTKNNLTSESSVISQTYNAQGNPIS----- 1671  
 552 QSTPARSALMRGLGVNSGELYQLGKLAGVLDAQ-NTITLSVFVSISSLYRITLLARVHLT 610  
 1672 -----ALHYVNGVNFNIAKALQOMALKLYPEIKKILGNDFSLSLNLKGDALNQLT 1724  
 611 -----VNELCMLYGLSPFNKGTASLSGELPRLVTLWLVQVQWLTEARITTEAI 660  
 1725 KLITPDSWKINEL-----IDNANNSVQNFNNG---TLIIGATKIGQTDNTNSAVVFGGL 1776  
 661 W-----LLCTPEFSGNISPEISNLNLRPSISDMAQSHNRELQAEILAPFAAT 711  
 1777 GYQPCDYTDIVC-QKFRG-----TYLQGLLSISADLGYIDTFNFAKEI-----YLTGT 1825  
 712 LHLASP-----DMARYIILLTDNLNRPGLGDIAGFMFTLVLBKESLANANETTOLVQF 760  
 1826 LGSGNAGTGGASVTFNSQTSLLINQANIVSSQTD--GIFSMLGQEGI--NKFVNAQGL 1881  
 761 CHVMAQL-----SLSVQTLRLSEAEISVLVISGFVAULCAKNOP-----AGQHNI 804  
 1882 ANILGEVAMQINSKAGGLGNLIVNTLG-----SDSVIGGYLTPEQKNQTLSQLGQNNF 1935

805 DTLFSLYRPHQWINGLGNPGSDTL--DMLRQQ-----TLTADRLASVNMGLDISMVTQAMVS 858  
 1936 DNL-----MNDGLNLTAKDLIRQKLGFWTGLVGLAGLIDQLQNEPKLIGS 1983  
 859 AGVNLQOCWODI-NTVLQWIDVASALHTMPSVIRTVNIRYVTAALKAESNLPBWDSEWQT 917  
 1984 MSINDLLSKGLFNQITGTFISANDIGQVISVMLQDIVK-----PS-----D 2024  
 918 LAENWEAGLSTQQ-AQTLADYTAERLSSVLCNNFLANIQPEGVSLHSRDDLLSYFLIDNQ 976  
 2025 ALKNDVAALGKQMGIEFGLQDTLNSLESJQN-----QQ 2058  
 977 VSSAIKTRTALAEATAGIQLYINRALNRIEPN-----ARADVSTROFPTD 1020  
 2059 IKSVIDKVLAAKGLGSI---YEQGLGLDILPNLKKGIFAPVGLSQVWKQKGFDFSNAGNV 2115  
 1021 WTVNRYSTWGG-----VSRLVYYPENVIDPQRIQOTRMMDELEENISQSLSRDTV 1073  
 2116 FVQNSTFENANGGTLSEFNAGNSLIFAGNNHIAFTNHSGLTNLLNSQVSNVNTMLNASN- 2174  
 1074 EDAPKTYLTRFETVADLKVVSAHDNVSNTGLTWV-  
 2175 -----GLK-INATNNNVSVSQG-NLFINASCVCQSDPTTASATNPCTTAQ 2217  
 1111 ---GQTRNLPEYWRNVDISRMQ-----AGELAANAWEKWKIDTAVNP-----YK 1154  
 2218 NNASSSNASNNAPALANNDESIVVTANGFNFSGNIYANGVVDVFSKIGSANVKNLVLYN 2277  
 1155 DAIRPVIFRERLHLIWEKBEVAKNGDTPVETVDRFTLKLAFLEHDSGWSAPWSDYDITTO 1214  
 2278 NA-----QFQANNLATISNOAVLEKNA-----SFVTNNLNIQAGFNNNATQK 2318  
 1215 VEAVTDKKPDERLALAASGFOGEDTLLVFVY--KTGKSYSDFGGSKNV----- 1262  
 2319 IEVL-----QNLVIASNA-----SLSTGYIGLEVGGALNNLGAHFNLENSQTPVNP 2366  
 1263 -----AGMTIYGDGSPKMENTALSRY-----SOLKNTFDIHH 1295  
 2367 IOVGIIINLNTQTPFMNVSVANGGTYTLKS---SRVIDYNINPNSLQSVLK-LYTLIN 2422  
 1296 TQGNDLVRKASYRPAQDFEVPASLMGSAIGDLSLTWENGNIPOITSKYSSDNLATLH 1355  
 2423 INGNHIEBKNG-----VLTYLQQRVLQDKGLL--SVALPNSNANSONNLTSLSVL 2472  
 1356 NAAFTVYDGSVNRNKQISAMK--LTGVDGKSOY-----GNAPIIANTVKHYGY 1405  
 2473 HNQIKMSY---GNKVMDFTPPTQDYVIGIQGQALNQIEAVGGNNAIKWLTSLMETKE 2529  
 1406 SDLGGPITVYNTKNTYIASVQGHLMNADYTRRLILTPVENNYIARLPFFPSPNTILNTV 1465  
 2530 NPLFAPVILENHSILNEILGVTQDLQON---TASLISNPNFRNATSLLEM-----ASYTQ 2580  
 1466 FTVGSNKTSDPK-----KCSYAVDGNNSQGFQIFSS-----YQSSG 1501  
 2581 QTSRLTKSLDFRAREGSESNFSERLLELKNRFSDPNPESEVFKYSQLSKHPPNNLMIQGVG 2640  
 1502 WLDDITGINNT-----DIKITWAGSKTHETA--SDHIASLPANSPDAMPYTFKP 1550  
 2641 GASPISGNGTLYGLNVGYDRLVKSVILGGVAVYGYGFGNGNIMHSIANNVDVGVYARAF 2700  
 1551 LEIDASSLAFTNNIAPLDIVFETKAKDGRVLGKIKQTLV--KRVNYNP-----E 1598  
 2701 LKRNFTLSAN-----ETVGNASHINSSNLSVLNQRYNYNTWTTSVNGNYGY 2750  
 1599 DILFIRLT-----HSGAQYMQLVG--YRIRLNTLLASQLVSRANTGIDTILTMETORLPE 1651  
 2751 DFMEKQSVVLKPOQVGLSYHFIGLSGKMGKMQNPAYQQFVMSHNSPESVLTLMNGLESR 2810  
 1652 PPLGEGEFANFVLPKYPDPAEHDGDERWFKIHGN---VGGNTGRQPYVSGMLSDTSETSM 1707  
 2811 KYFGKNSY---YFVTRILGRDLLIKAKGDNVVRFGENT---LLYRKGEIFNT----- 2857  
 1708 TLFVPAEAGYMH-----EGVRLGVGYQKI 1732

```

Db 2858 --FASVITGEMHRLWLMYVAGVGLKMGLOQDL 2890
RESULT 10
F82884
Hypothetical protein UU495 [imported] - Ureaplasma urealyticum
C:Species: Ureaplasma urealyticum
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82884
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Casseell, G.H.
submitted to GenBank, February 2000
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min
A:Reference number: A82870
A:Accession: F82884
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5005 <GLA>
A:Cross-references: UNIPARC:UPI0000110214; GB:AE002148; GB:AF222894; NID:G6899495; PIDN:
A:Experimental source: serovar 3; biovar 1
C:Genetics:
A:Gene: UU495
A:Genetic code: SGC3

Query Match 1.7%; Score 220; DB 2; Length 5005;
Best Local Similarity 17.7%; Pred. No. 0.023;
Matches 522; Conservative 425; Mismatches 1055; Indels 952; Gaps 142;

Qy 113 FSPAGVTELYEAKDLHFPSSAYHLDN-----RRPDLADLTLSQSNMDEISTLTLSNEL 168
Db 1376 FSASGVITNL-----LADHTYRLTSVSLKQKPOLANVINNN-----NNEI 1417

Qy 169 LLEHI--TRKTGGD--SDALMESLSTYRQAITDPYHPQYETIRQV-----I 210
Db 1418 LNNIEDNQKTHLTSKSVINEISNLADAYPKNNGIYDVKEFINIKNKKLNKKYKV 1477

Qy 211 MTHSTLSALSRNPVMOQEGASLAI-----LANISP-ELYNIUTEI-TEKNADALFA 264
Db 1478 VFEDNEHQLISTNDLLVNKLDQTNLSLQNFSSNLKPNHLYRLKVVYGEQNPDAINE 1537

Qy 265 Q-----NFSENITPENF-ASQSWIAKYG-----LELSEVQKYL 298
Db 1538 QKNHIALNPVLSNSSFSTPAKIKVSKNAIDVMOQALIKLILDDSDNQLHSGDBININ 1597

Qy 299 ---MLQN-----GYSDSTSAYDNISTGLVNVNESKLEAYKITRVKTDYDKNIN--YF 347
Db 1598 RIKQTQNIISTPATISDQKRYKCMATNLAGLDYEVSVTIKQKT-----KNVSPILP 1653

Qy 348 DL-----MYEGNQQPIRANFKVSRFEGATL----- 373
Db 1654 ELPSGPIGFYTLAPVFKITSLSMDFVYHENKNTADLKINLRI-ENIGASLFNKDIIKFIK 1712

Qy 374 RKNAGPS-GIVGSLGPIANTNFKSN-----YL-----SNISDSEYKN 411
Db 1713 RKNQDKQISFTHKVSANDANYEWEFKOLLNRREYTLERVVYLRKQFNQVNSSESY-- 1770

Qy 412 GKVIYAYRTSATSATNGGIFTPESEYPLTIFALKLNAIRLCITSLGSPNELQTVRS 471
Db 1771 -IDLWIYB-----NLNKTFLPTKPL----- 1791

Qy 472 DNAQGIINDSVLTKVFTYLFYSHRVALSFDDAQVLNGSVINQYADDDSVSHFNRLFTNPP 531
Db 1792 ---GIIGAPI-----KEISDNGAKVQLKPAIN---DFDDVLKENQTFK--- 1828

Qy 532 LKGKIFEADGNTVSDPDEEQSTPARSALMRGLGVNSGELQGLKGLAGVLAQNTITLSV 591
Db 1829 -----FNIQPNENN-----GANLNEISEHEGKEVEIIDGQK----- 1858

Qy 592 FVSSLYRLTLIARVHQLTVNELCMLY-----GLSPFNGKTTASLSSGELPRL 639
Db 1859 -----FFVANLNKIKVKEKYKLYIPENQDQVNGVYKINFK-----NDYKEPN 1904

Qy 640 VIWYQVQWLTAEITTEAIWLLCTPBFSGNISPEISNLLNLRPSISEDMAOSHREL 699

```

```

Db 1905 VVY-----DASVNTTQTYVF-TNKFA-----IASFSNNL-----TDVDVANKONTSI 1945
Qy 700 QAEILAPPIAATLHLASPDMARYI-----LLWTDNLRPGGLDIAGFM-TLVLKESINANE 753
Db 1946 NLDSEVETIQG-YHF-----KAKYISDNRRVVWTYI-PAPYINGNKNKIALNPFELNQO 1999
Qy 754 --TTQLVQFCHVMAQLSLSVQTLRLSEAEISLVLI-----SGFAVLGAKNQ 797
Db 2000 LISNRLYTF-----AALYYS-KDVNVNDENHANNVKNVNPQTISTKPSSTYVDLKAQNA 2054
Qy 798 PAGQ-----HNIDTLPFLYRPHQWINGLNGSPGDTLMDLRQOT-----LTADR----- 840
Db 2055 DENKITLALLHSNDQIFEDKRNHLKIAKISIDELDAHQIINSTHDDYDLTLEKENNEW 2114
Qy 841 LASVMGLDISMVTQAMVSA-----GVNQLQCMQDINTVLQWIDVAS-ALHTM 886
Db 2115 LLKTLQNLINKENTKYRVKVKWFTSKPSTIYGVNKNIVYPINNHNNTDLKLTLEKSTL 2174
Qy 887 PSVIRTLVN-----IRYVTALNKAESNLPW-----DEWQTL 918
Db 2175 NSVKPTKTNFNKNESIKLMVGFNKTGSSLENKYAKLVYKDNNNQOIESDPVLLNSDTTK 2234
Qy 919 AENMEAGLSTQOAOQLADYTAERLSSVL---CNWFLAN-----IQPEGV 959
Db 2235 PFNKEFSFNNQTSILVANRDPFVKLIISDTPADFNQANTLDTLPNFNKDDAKFSLEPTPI 2294
Qy 960 SL-----HSRDDLYSYFLIDNQVSAIKTTFLAEAI---AGIQLYINRNLNRIE-PNA 1008
Db 2295 SVNNVIOGSNTYDIBHLISFDYDDQDHNLDNDQITTYRKKGEQAWTSSKAGEVSXKH 2354
Qy 1009 RADVSTROFFDTVNNRNYSTGWGVSRL-VYYPENYIDPTQIGOTRMMDELLENISOK 1067
Db 2355 KINVTLK-----DLTPNTTYEI-GAISTHVSPPVKTSPIQVNETTKQLEKLD--LTK 2407
Qy 1068 LSRDTEVDAFKYTLTRETVDLKVVS-----AYHDNVNSNTGL-----TW 1109
Db 2408 VAHNFIXD-----FKLASNSKLNGVGYECADENKVFSGVLHKGHEGEYBIYAKF 2459
Qy 1110 VGQTRNLEPYWRNVDISRMQAGELAAANAKWTKIDTAVNPKYKDAIRPVIREFRLHLI 1169
Db 2460 VDKTNPKFA-----IEIKSSNNINSNQVYFELLKSKQITPN-----HQVELSEI 2506
Qy 1170 WVEKEVAKNGTDPVETVDRFTLKLAFLRHGDSWAPSMDYI---TTQVEAVTDKK----- 1222
Db 2507 CAIKNKGANNEIKKDLNENHLSLALGL-----APNKMVDQKPTITITIKYQDNQSVSA 2558
Qy 1223 -----PD-----TERLALAASGF-QGEDTLV-----FVYKTGKSYSDFGS 1258
Db 2559 HVKYIINDPDGILNSTCCKELLNVKALFNQTNQANLINNITIKYDQNKQFYCEF--- 2615
Qy 1259 NKNVAGMTIYGD-GSFK-----KMENTALSYSQKNTFDIHTQGNDLVRK----- 1304
Db 2616 --DIDNLTNQDYGIFEISFANKPIHAAFAKINDTQHPALLSTNNQDINKIALDCEVN 2673
Qy 1305 -----ASRFAQDPEVPASLNMGS--- 1323
Db 2674 KVTINKQDLDYDQNTKISWPTELSSNLNNKYIRVKYSVSGHKFQDDQFSASFATITNKK 2733
Qy 1324 ---AIGDLSLTM-----ENG-----NIPQITSKYSSNDLAIHLHNAFTVRY 1363
Db 2734 ASVEIGHDKLIANKAYEIQAIYFSDQNDNDFACTNQNQIPINSSITNKKIVINPSATTISS 2793
Qy 1364 DSGGNVRNKKQISAMKLTGVDGKSYGNAFIANTVHYGYSGL-----GGPITYVN 1416
Db 2794 HEVKNITSTADINCSISTNDGDIKPNQSIIG---HFAKGDDLTTKTFVANGVVANVN 2849
Qy 1417 KTKNYIASVQGHLMNADYTRRLILTPVE-----NNYARL---FEPPSPN-----T 1460
Db 2850 QTK-----LSPTLANLEPNTKYILVSVEFKQPPNAYKHNNKYNQDQFYSNDQTHSIIFT 2904
Qy 1461 ILNTVFTVGSNKTSDPKKCSYAVDGNNSQGFQIFSSYQSSGWLDDIDTGINNTDKITVMA 1520

```







Db 1483 -VNYLLDDIV---KIHSSSFL-----RIQLEQ-LSHVLVDKITAVLPKSLTIK 1526

RESULT 13

S59413

probable membrane protein YLR454w - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein L9122.5

C;Species: Saccharomyces cerevisiae

C;Date: 30-Nov-1995 #sequence\_revision 23-Feb-1996 #text\_change 09-Jul-2004

C;Accession: S59413

R;Kirsten, J.

submitted to the EMBL Data Library, March 1995

A;Description: The sequence of S. cerevisiae cosmid 9122.

A;Reference number: S59414

A;Accession: S59413

A;Molecule type: DNA

A;Residues: 1-2628 <KIR>

A;Cross-references: UNIPROT:Q06179; UNIPARC:UPI000006A416; EMBL:U22383; NID:G2264349; PI

A;Experimental source: strain S288C (AB972)

C;Genetics:

A;Gene: MIPS:YLR454w

A;Cross-references: SGD:S0004446

A;Map position: 12R

C;Keywords: transmembrane protein

F;7-23/Domain: transmembrane #status predicted <TM1>

F;173-189/Domain: transmembrane #status predicted <TM2>

F;835-851/Domain: transmembrane #status predicted <TM3>

F;2569-2585/Domain: transmembrane #status predicted <TM4>

Query Match 1.7%; Score 217; DB 2; Length 2628;

Best Local Similarity 17.9%; Pred. NO. 0.011;

Matches 511; Conservative 398; Mismatches 1000; Indels 940; Gaps 136;

Qy 144 DLADLTLSQSNWDTISLTLSNELLEHTRKTGSDSDALMESLSTYRQAIIDPYHQP- 202

Db 215 NLLMLFINKENV-----LMSNEKLLQRLADTTKANBELKDEDIAKMO--DLVYAMEK 266

Qy 203 ----YETIRQVIMTHDSTL----SALSRLPEYMGQAGASLAILAILANISPELYNLTIERIT 255

Db 267 FVDRIKPLKENNVTVKLIQKDFLTNHPPELLGMNKYISYNVLVSNL-----NFTNRF- 320

Qy 256 EKNADALPAQNPSENITPENFASQSWIAKYGYGLESEVQKYLGLMQLNGYSDSTAYVDNI 315

Db 321 -RNEMPGYTLIFEERDSFPKFS--IIMARF-----NI 349

Qy 316 STGLVNNESKLEAVKIRVTKDDYDKNINYFDLMYEGNQFFIRANFKVSRFEGATLRK 375

Db 350 YLNLRKHQSHAKQLKITEI-----PNVSI-----GETNLFQ-KEPLSNLHAKLE 397

Qy 376 NAGPSGIVGSLSGPLI----ANTWPKSNYLSNI----SDSRYKNGYKIYAYRTSSTSATN 428

Db 398 NA-IFNIKGNISLTIDMDPNVISPIKFLSNIKVFTSSCPKNKI----LKENSHVKFLT 452

Qy 429 QGGGIPTSPVLTIPALKNAIKARLCLTSLGSLPNELOTIVRSNDAQIINDS---VLTK 485

Db 453 RRRVLDYFKCFPLINKFT-----LDDPKFVINDKODLIIGK 491

Qy 486 --VFYTLFYSHRYALSPPDDAQVLNGSVINQYADDSVSHENLFWTPPLKGGKIPEADGNT 543

Db 492 FSVFMISHSKRYTL-----GNN 509

Qy 544 VSIDPDERQSTFARSALRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSIYRLTL 603

Db 510 LMEKEEFTQHIFYESHW-----NVELLDM-KLOHIKHQK-----YEHTIL 549

Qy 604 ARVHQLTVNELCMLYG--LSPNGK--TTASISGSELPLRLVTLVQVOTWMLTEAEITTEAI 660

Db 550 -RVDISIAEEKVQLLPDILCSANADITLMLDLSLPTVMVMSLSELVHLNLSQLANVEENY 608

Qy 661 WLLCTPESGNI---SPRISNLLNLRPSISEDMAQSHNRELOAEIL-APFIAATHLAS 716

Db 609 FKFEYEFASNLQNMKAECNNAKCLR-----QKEILSPDFMFOOL----- 649

Qy 717 PDMARYILLMTDNLRPOGLDIAG-----FMTVLVKESLNANETTQIV-----QFCHVM 764

Db 650 PDFDYIKI---NIR-----DISSTLGARSVPMPDRFSSVDSQSKDLIDGKLRKYCNTV 702

Qy 765 AQLSLSVQTLRLSEALSVLVISGFAVLGAKNQAGQHNIIDTFLSLYRHFHQMINGLGNPG 824

Db 703 EKLOI-----ALFGDKT-----QWNKI---G 721

Qy 825 SDTIDMLRQOQTLT-----ADRLASVMGLDISMVTQAMVSAGVNOLOQCQDINTVLQ 875

Db 722 SNHATMVRSGQLTFNSKDNKQPNHKSSIAIDDDISTSDA--TEVNHL--W-NINLLVN 775

Qy 876 WIDVASALHTWPSVIRTLNIRYITALKAESNLPSPDEWOTLAENMEAGLSTQQAOTLA 935

Db 776 --DITTSI-----IGETPEVSELS 793

Qy 936 DYTAERLSSVLNFWFLANIOPEGVSLHSRDLVSFVLIDN--QVSSAIKTKRLAEAIAGI 993

Db 794 TKTVSKVSNLSIKLP-----PDTSFSSNESDSKILIIQINHSRGTSVSVLSMISIFLAVSGI 848

Qy 994 QLYINRALNRIEPNARADVSTROPFTDWTNNRYS-----TWGGVSRVLVYPENYIDPTQ 1048

Db 849 HT-LNQIFGHCHQMRQSKTKQYFLALSESCKSKCIKWG----- 890

Qy 1049 RIGOTRMDELLE-NISQSKLSRDTVEDAFKTYL--TRPETVADLKVVSAYHDNVNSNTGL 1106

Db 891 -----QKELLEINFSEYISQII--ALPNGLRTKPEPTSTP--ITVKNCTISVSQ 939

Qy 1107 TWFGQTRENLPEVYWR-----NVDISMQAGELAANAKWKTKIDTAVNPYKD 1155

Db 940 YFRMVESPTQPNFWRMFCINGFKVMIHDLKQOMKKL--NSLQWELKPSAITLEND 997

Qy 1156 AIR-----PVIF--RERLHLIWEKEE--VAKNGTDPVETYDRFTLKL 1194

Db 998 SWHFSIPHPPEMKIIDSIPFIKSIQMLYSKATSKDDLIIFPHKIETPLSLPKLK- 1056

Qy 1195 AFLRHGWSWAPWSYDITQVEAVTKKPDTERLALASGFGQEDTLLVFVYTKGKSVS- 1253

Db 1057 -----SKRWLF-----SISDDPLEALNTIFIQIGLQERERLAKLQEFNKRLSE 1100

Qy 1254 DFGGSKNVAGMTIYGDSFKKMENTALSRYSQKNTFDIHTQGNIDLVRKASVYFAODF 1313

Db 1101 DLIKSQNAKEM-----KDDFEAIDNAI-----LKHRTGLWAKDGKKELRKS---ATDS 1146

Qy 1314 EV---PASLNNGSAIGD-----DSL----- 1330

Db 1147 EILPTPALNNGKRRDRPDRTQFISPEIENAYNTLLANFSDSWIKRVKVKVERREFD 1206

Qy 1331 -----TVMENGNIPOITSKYSSDNLAITLHN 1356

Db 1207 KNFSFLMGFIDYTKLPKDINKKVLFPSTNPFMLNLIENIDIDIIRPSCGIENIPNF1HD 1266

Qy 1357 AAFVTRVYDGSNVIRNQISAKMLTGVDGK-----SOYGNAFIIANTVKHYGYSDL 1408

Db 1267 V-----GKG--VPKNTSEYSIMI PMHLDAKSEVVRWHLRDYPLFVSIPLPSSTQSKETI 1318

Qy 1409 GGPITVYNTKKNYIASVQGHLMNADYTRRLILTP-----VEN--NYVARLEFPFPSPNT 1460

Db 1319 --PMRIYGD-----FMITEDMLQSDRELTLFVPLPSVTVENTDRYYSL-----FVPT 1366

Qy 1461 ILNT-VFT-----VGSNKTS-----DPKCSYAVDGNNSQGFQIF 1494

Db 1367 MTSAKIFITDLNFEINSNHTTRVTWGGSYQPAIQOTMQCLDNFSK--PPLDPSVLKGLFWDK 1424

Qy 1495 SSYQSSGWLDI-----DTGQNNIDKITWAGSKTHTFTAS--DHIASLPSANSPDAMPVT 1547

Db 1425 TRYLFHGKINIVWKKRGKFEISLGAKSPYMLGGSAGFIVGFDGNVNLKNE--DNDPK 1483

Qy 1548 FKPLEIDASSLAFTNNIAPLDIVFETKAKGRVLGK-----IKQTLVSXRV 1593

Db 1484 FLSCSADKVFHSIPNYFAKPLLVMSRPSNTMTFIPNQDDTNMQRYASFYLLNTTSSQNE 1543

Qy 1594 NYPNEDI--LFLRETH-----SGAQYWLGVYRIRLNTLLASQLVSRANTGIDTILT-METQ 1647



Db	1544	KADKEIMKGFIETGIKLSCGMLTDMGILFERLGPSSLNERTFESKXHYIURLCNPIYVQ	1603
Qy	1648	RLPEPPLGEGGFANFVLPKYDPAASHGDBR	1684
Db	1604	DLSKHDSYAGERSDFIHMSFGLSNSNSAYNAMQLSPNGKAFVFWKWSFGSFPVRRGP	1663
Qy	1685	-----VCGNTGROPYYSGMLSDTSETSMTLFVPIABGYMYHGEVLGVGYQKITYDNTW	1738
Db	1664	LFGLOSIPKPFGEHLTISYHADVS-----PLFINYM-----YHNADA-----DQILRKNYL	1710
Qy	1739	E-----SAFFYDETKQOQVFLINDADHDSGMTQOQIVKNIKYKGEF-----NVSIA	1785
Db	1711	EVAEFAGLKAKSSHFIMDLHORKEVL-----TEYQAGLVNRRVRVMKLFPLAGDVVQCDVDIR	1767
Qy	1786	TGYSAPMDENSASALYIYWEFLFYIYTPMFCQRLLEKQ-----PDEATOMIN-----Y	1832
Db	1768	T-----VSGEF-----SKLNY-----IEEKDAEYDIPDNMSWLDITDFODAF	1806
Qy	1833	VYNPAGYIVANGEIAPWJNCR-PLUEETTSWNANPLDAIDPAQANDPMHYKIATFMRLL	1891
Db	1807	FINPDNYLPKIKMPFAFSQFQAKRASCYDK--YQVDPKTKCPIPTFDNRVSHGCTLG	1864
Qy	1892	DQLIRGDMAYRELT-----RDALNEAKMYVVRTLELLGDEPEDEYGS-QQWAAPSLSGA-	1944
Db	1865	HNVSLRDLVEKRVTVLKKFREKLQEG-----IRKNSAGVSEENLNDLLSKANSSVENAE	1920
Qy	1945	-----ASQTVQAAVQODLTMLGRGVSKNLATNSLVGLFLPEYNPALTDY	1990
Db	1921	LLLKDFOKIPKQHEAGQTQEPFHFDPSLNL--RNTKTKLQFHRFFI-----NVLKWK	1973
Qy	1991	WQTLRLRLFNLRHMLSIDGQPLSLAI-----YAEPTDPKALLTSMVQASQGS	2038
Db	1974	NEDARSAIFKFFYYANLSNFASLASGKGLREFEDVIKQREMTDD-----TTSMEAIPG	2029
Qy	2039	AVLPGTSLSYRFPVMBLRTNLV-----AQITQFTSLLSMAEHDDADELTLLLLQOQWEL	2094
Db	2030	DKANTTKQCHSCDDTEPTENLLNIFEKNITQLSCDIKNKIHH-----KF	2074
Qy	2095	ATQSIRIQORTVEVDADIAVLAESRRSAQNRLE-----KYQQLYDEDI-----NHG-	2141
Db	2075	FVQFIPFQIQLTSLNPEACVLVSSPFPMUKLTLEFDANTTSNTYMQDIFLRHGILFGNA	2134
Qy	2142	-----PQRAMSLDDAAAGOSLAGQVLSIABGVADLPNVFGLACGSRGWAALRASASV	2195
Db	2135	NAFLFNKKDYQEPFELFYGSSSYGQ-----DKKEQWPPWLGLELGEF--PSALKKKAVV	2186
Qy	2196	MSLSAT-----ASQYSA--DKISRSFAYRRRRQEWEIFQRDNADGEVKQMDAQLS-	2243
Db	2187	RNISALLHHQKLAPFSKAKYDSLKDKI-----EDNICGYVPQVNVQVNSD	2230
Qy	2244	--LKIRREAAQOMVEYQETQQAHTQAQLELL-----QRKFTN-----	2278
Db	2231	EYLMLTQWALKLFL-VYEPEDEELKKYIEKLIIGYDIYDTAQRKTFVNDLHDSEQILAVV	2289
Qy	2279	-----KALYSMMRGKLSAIYYQFPLT-----QSFLMAQEAALRELTNDGVTPI-RG	2325
Db	2290	EKELLFKRSLDDI-GKL-----DLSNIHMERMQLLURLYLIRKVTFTSGNNYINRT	2340
Qy	2326	GAMNGTTAGLMAGETLLNLNAEMEKWMLERDERALEVTR-----TVSLAQFYQ	2373
Db	2341	LVWN-----IKVNETILHLLDKIDKFDLIAVAKLNFQIHTMGLRKNTVTVKQMQIPD	2395
Qy	2374	ALSSDNFN--LTEKLTQFLREGKNGVCSAGNELKSNRQIEASVRLSD-----LKIPSDYP	2427
Db	2396	IGENVNTHCJLGPLIT-----SSGND-----TVGLASDVPVLQVITWDVD	2434
Qy	2428	ESLGNTRQLQOVSVTLPALVGPVEDIR-----AVLNYGGSIVMPRGCSAI-----	2472
Db	2435	KPVGGIKVKNVETLLSSLIKLEEDRLNKLFEWLSLKELIYDGNDDDDGASIFDMWS	2494
Qy	2473	--ALSHGVNDSGQFMPLDFND-----SRYPFEGISVND-----SGSL-----	2507

[illegible]







```

Db 3103 -----NTGTGKGENASKVLPHKP--PIDTASMKAKAAKQNSIEMTG--SLOKSP--SABA 3150
QY 2058 RNLVAQ-LTOFGTSLLSMAEHDDADELT-TLLLOQSMELATOSIRIQRTVDEVDAIAV 2115
Db 3151 MGIYSQKTEGANSKFGIAQ--GAVETILTASAQSGVTCRDISV-----SNIGV 3197
QY 2116 LAESRRSAQNLEKYQOQLYDEINHGEOQAMSLDAAAGQSLAGQVLSIAE-----GVA 2169
Db 3198 ASETVTQPSNK-----ETGIGFGASNLIAPPESAETEFTGKISNLSQTSLNKMAAS 3249
QY 2170 DLVPMVFG-LACGSRWG--AALR--ASASVMSLSATASOVSADKISRSEAYRREROEWE 2224
Db 3250 DVGTTVESKIQAPGNYGDVSLQKVASDAILKAWQASRDSAISVD----FRQDRE--T 3303
QY 2225 IORDNADGEVKQMDAQLESKI---RREAAQMOV----EYQETOQA--HTQOQLEL----- 2271
Db 3304 VSAEKSDLNFKSTNSETQKLLFESKEBESGIFIRSSHEYETQKTLRHSASRESASRT 3363
QY 2272 -----LQRKFTNKALYSWMRGKLSAIYYQFFDLTOSFCLMAQOALRRELT-----DNG 2319
Db 3364 VTAPTNQEVQMNFDKKVEDSVAEGSLSIGIVR--ESSQSEVMQHAER-TSELTKLSMNEE 3420
QY 2320 VTFIR--CGAMNGTTAGLMAGETLLLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSS 2377
Db 3421 VAGVRAVSETTNETPRGYQQGDVQVOTGAAMGRIEAPRPOG----- 3462
QY 2378 DNFNLTEKLTQFLREGKGNVGASGNELKLSNRQIEASVRLSDLKIFSDYPESLGNTRQK 2437
Db 3463 -EAEITQKLRRTL-----SVERSASAKASMAESQTVTQIQKREDSLASEY--SVRDTLLK 3514
QY 2438 QVSVT 2442
Db 3515 SSSVS 3519

```

Search completed: February 16, 2006, 21:44:50  
 Job time : 89.6609 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 57.1714 Seconds  
(without alignments)  
3670.210 Million cell updates/sec

Title: US-10-754-115-34

Perfect score: 13043

Sequence: 1 MYSTAVLNKISPTRDGQTM.....KALLESLDIILHITYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_AA.\*  
1: /cgm2\_6/ptodata/1/iaa/5 COMB.pep.\*  
2: /cgm2\_6/ptodata/1/iaa/6 COMB.pep.\*  
3: /cgm2\_6/ptodata/1/iaa/H COMB.pep.\*  
4: /cgm2\_6/ptodata/1/iaa/pCTUS COMB.pep.\*  
5: /cgm2\_6/ptodata/1/iaa/RE COMB.pep.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5148.5	39.5	2522	US-09-251-645-13	Sequence 13, Appl
2	5043	38.7	2516	US-08-851-567B-47	Sequence 47, Appl
3	5043	38.7	2516	US-09-817-514A-2	Sequence 2, Appl
4	4985.5	38.2	2504	US-08-851-567B-12	Sequence 12, Appl
5	4985.5	38.2	2504	US-09-817-514A-8	Sequence 8, Appl
6	3303	25.3	1849	US-08-851-567B-49	Sequence 49, Appl
7	3278.5	25.1	1844	US-08-851-567B-53	Sequence 53, Appl
8	1632	12.5	573	US-08-851-567B-55	Sequence 55, Appl
9	1615	12.4	579	US-08-851-567B-51	Sequence 51, Appl
10	1591	12.2	1565	US-08-851-567B-59	Sequence 59, Appl
11	1553	11.9	1189	US-08-851-567B-26	Sequence 26, Appl
12	1401.5	10.7	845	US-08-851-567B-37	Sequence 37, Appl
13	981	7.5	562	US-08-851-567B-30	Sequence 30, Appl
14	593	4.5	627	US-08-851-567B-28	Sequence 28, Appl
15	429.5	3.3	1095	US-08-851-567B-34	Sequence 34, Appl
16	309	2.4	965	US-08-851-567B-57	Sequence 57, Appl
17	254.5	2.0	10182	US-09-134-001C-3159	Sequence 3159, Ap
18	248.5	1.9	603	US-08-851-567B-35	Sequence 35, Appl
19	246	1.9	5024	US-09-710-279-2964	Sequence 2964, Ap
20	223.5	1.8	2383	US-09-492-709A-302	Sequence 302, App
21	223	1.7	2710	US-08-480-604A-6	Sequence 6, Appl
22	223	1.7	2710	US-08-405-496A-6	Sequence 6, Appl
23	223	1.7	2710	US-08-915-136-6	Sequence 6, Appl
24	223	1.7	2710	US-08-957-310-6	Sequence 6, Appl
25	223	1.7	2710	US-10-011-366-6	Sequence 6, Appl
26	223	1.7	2710	US-09-084-517-6	Sequence 6, Appl
27	200	1.5	2188	US-09-328-352-763	Sequence 763, Ap

#### ALIGNMENTS

##### RESULT 1

US-09-251-645-13  
; Sequence 13, Application US/09251645  
; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.

; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: CGC1963/A

; CURRENT APPLICATION NUMBER: US/09/251,645

; CURRENT FILING DATE: 1999-02-17

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 13

; LENGTH: 2522

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-09-251-645-13

Query Match 39.5%; Score 5148.5; DB 2; Length 2522;

Best Local Similarity 43.0%; Pred. No. 0;

Matches 1121; Conservative 464; Mismatches 822; Indels 197; Gaps 55;

Qy	22	LADLYLSFSLRKIFDDQLSGEARHLYHTIEQKNRLLERARIFTRANPOLSGAIRL	81
Db	28	LTDSHSFNFHQCVSEHLSSEADLYHQAQKNRLYEARILKRTNPQLQNAVHL	87
Qy	82	GIERDSVSR-SYDWMFGARSSFFVKPGSVASMFSPAGYLTLYREAKDLHFSSAYHLDN	140
Db	88	AIVAPNAELICYNQFSGRASQYVAPGVTSVMFSPAAYLTLYREARNLHASDSVYRLDT	147
Qy	141	RRPLADLTLSQNMDETSITLSNELLLEHITKTKGDS-DALMESLSTYRQAITDTPY	199
Db	148	RRPLKGNALSQNMDETSITLSNELLLEHITKTKGDS-DALMESLSTYRQAITDTPY	207
Qy	200	HOPVETROVTMTDSTLSALSRNPVQVQAGSLAILANISPELYNLTETEETKNA	259
Db	208	HDATENVKVLQDPGLEQLNASPAIAGLMHQSLGINASISPELYNLTETEETKNA	267
Qy	260	DALFAQNFSENIPTENFASQSWIARYKYLELSEVQKYLGMQLQ-CYSDSTSAVDNI	318
Db	268	EELYKKNFG-NIEPASLAMPYLRYYNLSDELSQFQKASNFQOQSVSN---NQLITP	323
Qy	319	LVVNESKLEAYKTRVKTDDYDKNINYFD--LMYEGNNQFPFIRANFKVSREFGATLRKN	376

28	199.5	1.5	3290	2	US-09-328-352-5486	Sequence 5486, Ap
29	192.5	1.5	1619	2	US-09-328-352-7347	Sequence 7347, Ap
30	192	1.5	2301	2	US-09-822-871-4	Sequence 4, Appl
31	191.5	1.5	2039	2	US-09-077-098A-7	Sequence 7, Appl
32	191.5	1.5	2039	2	US-10-192-584-7	Sequence 7, Appl
33	185.5	1.4	2285	2	US-09-308-375-2	Sequence 2, Appl
34	185.5	1.4	2285	2	US-09-932-183A-2	Sequence 2, Appl
35	185.5	1.4	4536	2	US-09-180-422B-27	Sequence 27, Appl
36	185.5	1.4	4536	2	US-09-108-006C-1	Sequence 1, Appl
37	184	1.4	2172	2	US-09-583-110-5266	Sequence 5266, Ap
38	184	1.4	2238	2	US-09-107-433-4506	Sequence 4506, Ap
39	184	1.4	2777	2	US-09-543-681A-6124	Sequence 6124, Ap
40	182.5	1.4	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	182.5	1.4	2366	1	US-08-405-496A-10	Sequence 10, Appl
42	182.5	1.4	2366	2	US-08-915-136-10	Sequence 10, Appl
43	182.5	1.4	2366	2	US-08-957-310-10	Sequence 10, Appl
44	182.5	1.4	2366	2	US-10-011-366-10	Sequence 10, Appl
45	182.5	1.4	2366	2	US-09-084-517-10	Sequence 10, Appl











```

Qy 2224 EIQRNADGEVQKMDAQLESKIRREAAQVVEYQETQQAHTQAOLELORKEFTNKALYS 2293
Db 2193 EIQRNAAELKQDAQLESKIRREAAVAVLQKTSKQQEQTSQALPQLOKFSNQALYN 2252
Qy 2284 WMRGKLSIYYQFFDLTQFCLMAQEAALRRRLTONGVTFIRGGAWNGTTAGLMAGETILL 2343
Db 2253 WLRGSLAIIYQFYDLAVARCLMAEQAYRWELNDDSAFFIKPGAWQGTYAGLLAGETIML 2312
Qy 2344 NLAWEKVMWRDRERALEVTRTVSLAQFYQALSSDN--FNUTEKLTQFLREGKNGVQASG 2401
Db 2313 SLAQMEDAHLKRDRALEVERTVSLAEVYAGLPKONGPFSLAQBIKIDKLVSQSGSAGSN 2372
Qy 2402 NELKL-----SNRQIEASVRLSDIKIPSDYPESIGNTRQLKOVSVTLPALVGPVEDIRAV 2456
Db 2373 NNLAFAGAGTDTKTSLOQASVSADLKIRDYFASLQKIRRIKQISVTLPALLGPPYQDVQAI 2432
Qy 2457 LNYGGSIVMPGCGSAIALSHGVNDSGQFMLDFNDSRVLPPFGISVNDSGSLTSLFFPDAT- 2515
Db 2433 LSYDGKAGLANGCEALAVSHGVNDSGQFQDFNDGKFLPFEGIAI-DOGTILTSLFFPNASM 2491
Qy 2516 ---DRQKALLESLSDIILHIRYTR 2537
Db 2492 PEKGQATMLKTLNDIILHIRYTIK 2516

RESULT 4
US-08-851-567B-12
; Sequence 12, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merio, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851.567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255

```

```

; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-851-567B-12

```

```

Query Match 38.2%; Score 4986.5; DB 2; Length 2504;
Best Local Similarity 42.3%; Pred. No. 0;
Matches 1110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;

Qy 9 NKISPTRDG-----QWTMLADQLVSPSELKRPDDQLSWGEARHLHYETIEQKKNRL 63
Db 3 NLSLSTDTTCQKQLTFCRAEIALYPDTPFREKTRGMVNWGEAKRIYIEIAQAEQDRNLLH 62
Qy 64 EARIFTRANPOLSGAIRLGIERD-SVSRSYDEMFGARSSSVKVPKPGSVASMFSPAGYLTEL 122
Db 63 EKRIFAYANPELLKNAVRLGTQMLGFTQGSIDFLGNRADYNAAPGVSASFSPAYITEL 122
Qy 123 YREAKDLHFSSAYHLNRRPDLADLTLSQSNMDTEISTLTLSNELLEHITRTKGDS 182
Db 123 YREAKNLHDSSIIYLDKRRPDLASMLSQKNDEEISTLALSNECLAGIETTKGSD 182
Qy 183 ALMESLSTYQADTTPHQPYETIRQVIMTHDSTLSALSRNPVPMVGAGASILLAIANI 242
Db 183 EVMNMLSTYRLSGETPYHHAYETVREIVHERDPGRHLSQAPIVAAKLDPTVLLGISHI 242
Qy 243 SPELYNLTBEITEKN---ADALFAQNFSENIIPENFASQSIWAKIYGLSEVQKYLGM 299
Db 243 SPELYNLTBEITEKN---ADALFAQNFSENIIPENFASQSIWAKIYGLSEVQKYLGM 299
Qy 300 LQN-GYSDSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINIFDLMEYGNQOFF 358
Db 302 LSHVGYG-----SDILVPLVDGVGKMEVVRVTRTPSDNYTSQTYIELYPQGDNYL 354
Qy 359 IRANFKVSRFPGAT---LRKNAGPSGIVGSLGP---LIANTNPKSNYLSNISSEYKNG 412
Db 355 IKYN--LSNSFGLDDFYLYQKQSGADWTEIAHNPYPDMVINQKYESQATIKRSDS--NI 410
Qy 413 VKIVAVRYTSTGATNQGIGFTFESYPLTIFALKLNKAIKRLCLTSLGSLNELQTVRS 472
Db 411 LSLGLQWHS--GSYNFAANFKIDQSPKAFLLKMKKALRLKATGLSATLERIVDSV 468
Qy 473 NAOQIINDSVLTQVFTLYSHRYALSFDDAQVINGSVINQYADDDSVSHFNRLFNTPPL 532
Db 469 NSTKSIITVEVLNKVYRVKPYIDRYGISEETAAILANINISQAAVGNQLSQPEQLFNHPL 528
Qy 533 KGKIFE-ADGNT-----VSIDPDEQSTFASALMRGLGVNSGELYOL-----GKLAG 579
Db 529 NGIRYEISEDNSKHLPNPDLNLKPDSTGDDQKAVLKRAFOVNASELYQMLLITDRKDG 588
Qy 580 VL--DAQNTITLSVFVITSSLYRLTLARVHOLTAVNELCMLYGLSPFNKKTASLSGELP 637
Db 589 VIKNNLEN-----LSDLYLSLLAQINFLIAELNILLVICGYGDTNIIYQITDNL 640
Qy 638 RLVIWLYQVTOWLTEABITTEAILWLCTPFGSNGISPEISNL-----LNNURPSISED 690
Db 641 KIVETLLWITQWLKTQKVTVDLFLMTATYSTTLTPEISNLTATLSSTLHGKESLIGED 700

```



; SOFTWARE: PatentIn version 3.0		; SEQ ID NO 8		; Score 4986.5; DB 2; Length 2504;	
; LENGTH: 2504		; TYPE: PRT		; Pred. No. 0;	
; ORGANISM: Photorhabdus luminescens		; Mismatches 434; Indels 217; Gaps 48;		; Matches 1110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;	
US-09-817-514A-8					
Qy	9	NKISPTDQ----	QMTLADLOVLSFSELRKIFDDOLSWGEARHLHYETIEQKNNRLL	63	
Db	3	NSLSTIDTICOKLQUTCPAEIALYPTDTFREKTRGMVNGEAKRIYELIAQEQDRMLH	62		
Qy	64	EARIFTRANPOLSGAIRGIERD--SVRSYDEMFGARSSFPKPGSVASMPSPAGYITEL	122		
Db	63	EKRIFAYANPLLNKNAVRLGTQMLGF1QGYSDLFGNADNYAAPGVSASMPSPAAYITEL	122		
Qy	123	YREAKDLFPSSAYHLNRRDPLADLTLQSNDMTEISTLTLNELLLEHITRTKGGSD	182		
Db	123	YREAKNLHDSSTIYLLDKRRPDLASLMLSQKNMDEEISTLALSNECLAGIETKTKSQD	182		
Qy	183	ALMESLSTYRQAITDPYHQPETIRQVIMTHDSTLSALSRNPVYMGQAGASLAILANI	242		
Db	183	EVMDMLSTYLSGETPYHAYETVREIVHERDPGRHLSQLAPIVAAKLDPVLLIGISSHI	242		
Qy	243	SPELYNLTBEITENK---ADALFAQNFPSENIPTENFASQSWIAKYTGLELSEVOKYLM	299		
Db	243	SPELYNLTBEITENK---ADALFAQNFPSENIPTENFASQSWIAKYTGLELSEVOKYLM	299		
Qy	300	LON-GYSDSTISAYDNIISTGLVANNESKLEAYKITRVKTDYDKNINYPDLMYEGNNOFF	358		
Db	302	LSHVGS-----SDILVPLVDGVGKQEVVTRTPSDNYTSGTYIELYPOGGDYL	354		
Qy	359	IRANFKVSRFGAT---LRKNAGPSGIVGSLGP---LIANTNPKSNYLSNISDSSEYKNG	412		
Db	355	IKYN--LSNSELDDFYLQYKDGSGADWTEIAHNPYDPVINOQYESQATIKRSDS--NI	410		
Qy	413	VKIYAYRTSTSNQGGIFTFESYPLTIFALKANKAIRLCUTSGLSNPELQTVRS	472		
Db	411	LSIGLQRWHS--GSYNFAAANFKIDQSPKAFLLKMKNAIRLLKATGLSFATLERIVDS	468		
Qy	473	NAQGIINDSVLTVPYTYLFSHYRALSFDQAQVNGSVINOYADDSDVSHFNRLFTNPPL	532		
Db	469	NSTKSIITVEVLNRYKVPYIDRYGISEETAAILANINISQAQVNGSQLSQFEQFNHPP	528		
Qy	533	KGKIFE-ADGNT-----VSIDPEEQSTPARSALMRGLVNGSGELYQL-----GKLAG	579		
Db	529	NGIRYEISEDNKSLPNPDLNLKPDSTGDDQKQKAVLKRAFOVNASELYQMLLITDRKEDG	588		
Qy	580	VL--DAQNTITLVSFVSSLRLLRLARVHOLTWNELCMLYGLSPFNGKTTASLSSGELP	637		
Db	589	VIKNNLEN-----LSDLYLSLLAQIHNLTIAELNILLVLCGYGDTNITYQITDNL	640		
Qy	638	RLVTLVQVOTWLTAEAITTEAILLCTPESGNISPEISNL-----LNNLRPSISED	690		
Db	641	KIVETLLWITQWLTKQKWTVDLFWMTATYTTLTTPESINLTATLSTLHGKESLIGED	700		
Qy	691	MAQSHNRELQAEIILAPFAATLHLASPDMAFYIILLWTDNLPPGGLDIAFGMTLVLKESLN	750		
Db	701	LKRA-----MAPCFTSALHLSQEVAYDLLLLWIDQIQPAQITVDGFW-----EVQ	746		
Qy	751	ANETT-QLVQPCVMAQLSLSVQTLRLSEALSVLIVISGFAVLGAKNQPAQCHNIDITFS	809		
Db	747	TTPTSLKVTITPAQLAQLSLIYRIGLSETLSLIVTQSSILLVAGKS--ILDHGLLTLMA	804		
Qy	810	LYRHOWINGNFGNSDPLDMLRQOTLTADRLASVMGLDLSMTQA--MVSAGVNOQCW	867		
Db	805	LEGFHTVWNGJGQASLILAAKQCALITVTVQAQMNKEESLLQVAAQVKEKDIKLTISW	864		
Qy	868	QDINTVLQWIDVASALHTMPSVIRTFLNRYVTVALNKAESNLPSWDENQTLTAENNEAGLS	927		

865	TQIDALLQWLQWSSALAVSPDLDLAGMALKVGDHNYA-----AQWAAAAALMAD-H	915	
Qy	928	TQQAQTLADYTAERLSSVLQWFLANLQPEGVSLHSRDDLSYSLFDLNQVSSAIKTRTLA	987
Db	916	ANQAQKLD---ETFSKALCNYYINAVVDAAAGVDRNGLYTLLLDNQSADVITISRIA	972
Qy	988	EAIAGIQLYINRALNRJEPNARADVSTROFTDW-TVNNRYSTWGGVSRSLVYYPENYIDP	1046
Db	973	EAIAGIQLYINRALNRJEPNARADVSTROFTDWERNKYKSYSTWAGVSELVYYPENYIDP	1032
Qy	1047	TORIGOTRMMDLLENISQSKLSRDYVEDAPKTYLRTFETVADLKVVSAVHDNNVNTGL	1106
Db	1033	TORIGOTRMMDLLENISQSKLSRDYVEDAPKTYLRTFETVADLKVVSAVHDNNVNTGL	1092
Qy	1107	TWVVCOTRENLPYVYWRNVDIRMQAGELAAWKEWTKTDVAVNPKDAIRPVIPIERL	1166
Db	1093	TYFGIDQAAPGTWYKRSVDHSCCKENGKFAANAWGEWKNKTCVNPWKNIIIRPVVMSRL	1152
Qy	1167	HLIWEKEEVARNGTDPVETVDRFTLKLALFLRHDSWSAPWSYDITQVEAVTDKKPDTE	1226
Db	1153	YLLWLEQ--SKSDDGKTTIYQYNLKLALHRIYDGSWNTPTFDVTEKVKNYTSTDAAE	1210
Qy	1227	RLALAASQFQEDTLLVFPYKTKSYSDPFGSSKNQVAGMTIYDGGSPKMENTALSRY-S	1285
Db	1211	SLGLYCTCYQGEDTLLVMFYSMQSSSYTYDNNAPVTLGLYIFADMSDNNMTAQAITYWN	1270
Qy	1286	QLKNTFDIHTHQND----LVKASVRFACQDPEVPASL--NMGSATIGDDSLTWMENGNI	1338
Db	1271	NSYPOFTVWADPPSDKKVITRVRNRYAEDEIYPSVTSNSNYSGEDHSLTLMYGGSV	1330
Qy	1339	POITSKYSSDNLAILTHNAAPTFRYDGSNVIRNKQISAMKLGTVGDKSOGYGNAFIANT	1398
Db	1331	PNITFESAEALRLST-NMALSIIHNGYAGT-RRIOCNLMK-----	1369
Qy	1399	VKHGYSGLGSPITVYKTKYVYASVQGHLMADYTRRLILTPVENNYARLFEPPSP	1458
Db	1370	-----QYASLGDGKFIYDS-----SPDDANRPNLVP-----LKFQKDE	1403
Qy	1459	NT-----ILNT-----VFTVGSNKTSDFKKCSYAVD-GNNSQGFQ-----	1492
Db	1404	NSDDSIYENPSSSEDKKWFSSKDDNKTDYNGGTQCDAGTSNKDFYNNLQIEVIS	1463
Qy	1493	-----IFSSYQSGWMLDITGINNTDIKITWAGSKTHFTA--SDHTASLPANSFAMPY	1546
Db	1464	VTGGYWSYKISNPININTGIDSAAKVTVKAGGDDQIFTDADNSTYVPOQAPSPFEMIY	1523
Qy	1547	TFKPLEIDASSLAFNNIAPLDIVFETKADQGVLGKIKOTLSVKRVNYPEDILFLRET	1606
Db	1524	QPNULTIDCKNLNFIDNOAHIEIDFTAQDGRFLGAETFIIPVTXKVLGTENVIALYSE	1583
Qy	1607	HSGAQYMOGLVYRIRLNTLASQLVSURANTGIDITLMTWETORLPEPPLGGGFANFVLPK	1666
Db	1584	NNGVQYMOIGAYRTRLNTLPAQLVSRANRGIDAVLSNETQIQEPQLGAGTYVQLVLDK	1643
Qy	1667	YDPAEHGDERFKLHIHGNVGGTGRQPYSGMLSDTSETSWTLFPVPV--AEGYVMEHV	1723
Db	1644	YDESHTGNKFAIEYVDIFKENDSFVIYQEGELSETQTVVKVFLSFIETATGNKHLW	1703
Qy	1724	RLGVGYQKITVNTWESAFFYDETKQO----FVLINDADHSDGWTQOQIVKIKYKGF	1779
Db	1704	R--AKYQKETTDK-----ILFDRDTDEKDPHGWFSLDDHKTFSGLSAQALKN-----	1748
Qy	1780	LNVSITATYGAIPMDPNSSALYIYVELFYITPMCCFQLLOKQFDEATQWIVNYYPAGY	1839
Db	1749	-----DSEPMDFSGANALYFVELFYITPMMAHRLQEQNFDAANHFVYVWSPGY	1800
Qy	1840	IVNGEIAPIWNCRPLEETTSWANPLDAIDPDVAQNDPMHYKIATFMRLLDQILLRGD	1899
Db	1801	IVDGKIALYHNNVRPLEEDTSSNAQQLDSTDPDVAQDDPMHYKVATFMATDILLMARGD	1860
Qy	1900	MAYRELTDALNEAKQWTVRTLLELLGDEPEDYGSQQWAAFPSLSGAASTQVQAAVQODLTM	1959
Db	1861	AAVROLEEDTLAEAKQWTVTOALNLLGDEPQVMLSTWANTFLGNAASKTTQVROVQLTQ	1920

Db	865	QTDAILLQWLQMSALAVSPDLAAGMMALKYIGIDHNYA-----AQWAAAAALMAD-H	915		
Qy	928	TQQAOTLADYTAERLSSVLNCWFLANIQPSGVSLHSRDDLYSFLIDNOYSSAIKTTRLA	987		
Db	916	ANQAQKLD---ETFSKALCNYYINAVVDSAAQVDRDRNGLYTYLLLIQNSADVITSRIA	972		
Qy	988	EATAGIOLYINRALNRIEPNARADVSTRQFFTDW--TVNNRYSTWGGVSRLYVYPENYIDP	1046		
Db	973	EATAGIOLYINRALNRIEPNARADVSTRQFFTDW--TVNNRYSTWGGVSRLYVYPENYIDP	1032		
Qy	1047	TORIGOTRMDDELLENISQSLSRDVTVEDAFKTYLTFRTFVADLKVVSAHDNVSNTGL	1106		
Db	1033	TORIGOTRMDDELLENISQSLSRDVTVEDAFKTYLTFRTFVADLKVVSAHDNVSNTGL	1092		
Qy	1107	TFWVGQTRNLPYEWYENVDI SRMQACELAAWKEWTKIDTAVNPYKDAIRPVI FRERL	1166		
Db	1093	TYFIGIDQAAAPGTYYWRSVDHSCENGKFAANAGEWKNITCAVNPWKNIRPVVYSRL	1152		
Qy	1167	HLIWEKEEVAKNGTDPVETVDRFTLKLAFRLHDGWSAPWSYDITTTQVEAVTDKKPDE	1226		
Db	1153	YLLWLEQQ---SKSDDGKTTIYOYNLKLARIYDGSWNTPTFDVTEKVKNYTSTDAAE	1210		
Qy	1227	RLAALASFGQEDTLLVFKYTKGYSDFGSGNKNVAGMTIYGDGSPKMENTALSRY-S	1285		
Db	1211	SLGLYCTGYQGEDTLLVFMFYSMOSSSYSDTNNAPVTGLYIFADMSDDMTNAQATYWN	1270		
Qy	1286	QLKNTPTDI IHTQGN-----LVKASVYFAQDPEVPASL--NMSSAIGDDSLTWENGI	1338		
Db	1271	NSYPOFTVMADPSDNKKVITRVNNRYAEDYBPSVTSNYSNGWDSHSLTMLYGGSV	1330		
Qy	1339	POITSKYSSDNLAITLHNAFTVRYDGSNVRNKQISAMKLTGVDGSKSYGNAFIANT	1398		
Db	1331	PNITFESAEDLRLST-NWALSIIHNGYAGT-RIIQCNLK-----	1369		
Qy	1399	VKHVGGYDGLGPIITVNTKNYIASVQGHLMNADYTRRLILTPVENNYARLPEFPSP	1458		
Db	1370	-----QVASLGDKPIIYDS-----SPDDANRFLNLP-----LPFGKDE	1403		
Qy	1459	NT-----ILNT-----VFTVGSNKTSDPKKSVAVD-GNNSGQFQ-----	1492		
Db	1404	NSDDSI CYNENPSSDKWYFSSKDNKTADYNGTQCIDAGTSNKFYNYLQIEVIS	1463		
Qy	1493	-----IFSSYSSGSGWLDITDGTINNTDIKITVMAGSKTHFTTA--SDHIASLPANSFDAMPY	1546		
Db	1464	VTGGYWSYKISNPININTGIDSAKVTVKACAGDDQIFTDADNSTYVYVQQPAPSEEMIV	1523		
Qy	1547	TFKPLEIDASLAPTNNAIPLDIVFETKADQGRVLGKIKQTLVKRVNYPNEDILFLRET	1606		
Db	1524	QFNLTIDCKNLNFIIDQAHIEIDFTATAQDGRFLGAEFTIIPVTKVLGTENYALYSE	1583		
Qy	1607	HSGAQMOLGVYRIRLNTLLASQLSVSRANTGIDTILTWETORLPEPLGEGFFANFVLPK	1666		
Db	1584	NGVQYMQIGAYRTRLNLFQAQLVSRANRIGIDAVLSMETQNIQEPQAGTYVQLVLDK	1643		
Qy	1667	YDPAEHGDERWFKIHI GNVGNTGRQPYSGMLSDTSETSMTLFPVY---AEGYMHGCV	1723		
Db	1644	YDESIHGTNKSFALEYVDI PKENDSFVIYOGELSETSTQTVVVKVFLSFIEATGNKNHLW	1703		
Qy	1724	RLGVGYKIKIYDNTWESAFYFDETKQO-----FVLINDADHDSGMTQOQIVKNIKKYKF	1779		
Db	1704	R--AKYQKETTCK-----ILFDRTEDEKPHGWFSLDDHKTFFSGLSSAQAALKN	1748		
Qy	1780	LNVSIATGYSAPMDFNSASALYWFELFYTPMCMCFORLLQSKQFDEATQWNYNVNPGY	1839		
Db	1749	-----DSEPMDFSGANALYFWELFYTPMCMCFORLLQSKQFDEATQWNYNVNPGY	1800		
Qy	1840	IVNGEIPWLNCRPLEETTSWNAIPLDAIPDVAQNDPMHYKIATFMRLDQLIIRGD	1899		
Db	1801	IVDCKIAIYHWNVRPLEEDTSWNAQQLDSTDEDAVQDDPMHYKVATFMATLDMARGD	1860		
Qy	1900	MAYRELFDALNEAKMYVTRLELIGDEPEYGSQWNAAPSLSGNASQVQAAYQOQDITM	1959		
Db	1861	AAYRQLERDITLAEAKMYVTRLELIGDEPEYGSQWNAAPSLSGNASQVQAAYQOQDITM	1920		

QY 1960 LGRGVSKN--LRTANSVLGLFPEYNPALTDYQWTLRLRLNLRNLSIDQPLSLAIY 2017  
 Db 1921 LRLSRVKTPLLTGANSALTALFLPQENSKLKGYNRTLAQRMFNLRNLSIDQPLSLPLY 1980  
 QY 2018 ABPTDPKALLTSWQASOGGSNAVLPGTSLTRFPWMLERTRNLVAQLTQFGTSLLSMAEH 2077  
 Db 1981 AKPADPKALLSAVSGGADLPKAPLTHRRFPQWLEGARGLNQLIQFGSSLLGYSER 2040  
 QY 2078 DDADBLTLLLOQGMELATQSIROQRTVDEVDADIIVLAESRRSAQNRLEKYQOLYD 2137  
 Db 2041 QDAEAMSQLQTOASELILTSIRMQDNQLAELDSKTAQLVSLAGVQQRFPDSYSLYEN 2100  
 QY 2138 INHGQRAMSLDAAAGSLAGQVLISIAEGVADLVNPNVFLGACGSRWGAALRASAYMS 2197  
 Db 2101 INAGEORALALRSESAIESQGAQISRMAGAGVDMAPNIFGLADGGMHYGAATAYADGIE 2160  
 QY 2198 LSATASQYSADKISRSEAYRRRQWEIORDNADGEVQMDAQLSIRREAAQMVY 2257  
 Db 2161 LSASAKMVDKVAQSEIYRRRQEWKIQRDNQAQAEINQLNAQLESLSIRREAAQMVY 2220  
 QY 2258 QETOQANTQAOELQORFTNKALYSWMRGKLSAIYYOFFDLTQSFCLMAQELRRLTD 2317  
 Db 2221 LKTOQAQAQALTLRSFESQALYSMLRGLSGIYFQYDLAVSRCLMAEQSIQWEAND 2280  
 QY 2318 NGVTFIRGANGTTAGLMAGETLLNLAEMKWLDERDALEVRTRVTSLAQFYQAL-S 2376  
 Db 2281 NSISVPKGMQGTVAGLLCGEALQNLQAEMEEAYLKWESRALEVERTVSLAVVYDSLEG 2340  
 QY 2377 SDNFNLTEKLTQFLREGKNGVAGSNELKSNRQIEASVRLSDLKIPSDYPESL---CNT 2433  
 Db 2341 NDRFNLAEQIPALLDKGEGTAGTKENGISLANAILSASVGLSDLKIGTDYDPSVGSNKV 2400  
 QY 2434 ROLKQVSVTLPALVGPYDIDRAVLNMGYSVWPRGCSAIALSHGVNDGQFMLDFNDSKY 2493  
 Db 2401 RRIKQISVSLPALVGPYDQVQMLSYGSGSTQLPKGCALAVSHGNTDNGSQQLDFNDGKY 2460  
 QY 2494 LPFEGISVNDGSLTSPFDDTRDQKALLESIDILHRYTIR 2537  
 Db 2461 LPFEGIALDDQGTNLNLPFNATDKQKAILQTMDSIILHRYTIR 2504

RESULT 6

US-08-851-567B-49  
 ; Sequence 49, Application US/08851567B  
 ; Patent No. 6528484

GENERAL INFORMATION:

APPLICANT: Ensign, Jerald C  
 APPLICANT: Bowen, David J  
 APPLICANT: Petell, James  
 APPLICANT: Patig, Raymond  
 APPLICANT: Schoonover, Sue  
 APPLICANT: French-Constant, Richard  
 APPLICANT: Rocheleau, Thomas A.  
 APPLICANT: Blackburn, Michael B.  
 APPLICANT: Hey, Timothy D.  
 APPLICANT: Merlo, Donald J.  
 APPLICANT: Orr, Gregory L.  
 APPLICANT: Roberts, Jean L.  
 APPLICANT: Strickland, James A.  
 APPLICANT: Guo, Lining  
 APPLICANT: Ciche, Todd A.  
 APPLICANT: Sukhapinda, Kitisiri

TITLE OF INVENTION: Insecticidal Protein Toxins From Photobabidus

NUMBER OF SEQUENCES: 88

CORRESPONDENCE ADDRESS:

ADDRESS: Dow Agrosciences Patent Department  
 STREET: 9330 Zionsville Road  
 CITY: Indianapolis  
 STATE: IN  
 COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/851,567B  
 FILING DATE: 05-MAY-1997  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/063,615  
 FILING DATE: 18-MAY-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/395,497  
 FILING DATE: 28-FEB-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 60/007,255  
 FILING DATE: 06-NOV-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/608,423  
 FILING DATE: 28-FEB-1996  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/705,484  
 FILING DATE: 28-AUG-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seay, Nicholas J  
 REGISTRATION NUMBER: 27386  
 REFERENCE/POCKET NUMBER: 960296.93804  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 608-251-5000  
 TELEFAX: 608-251-9166  
 INFORMATION FOR SEQ ID NO: 49:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1849 amino acids  
 TYPE: amino acids  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

US-08-851-567B-49

Query Match 25.3%; Score 3303; DB 2; Length 1849;  
 Best Local Similarity 39.3%; Pred. No. 1.2e-252;  
 Matches 764; Conservative 339; Mismatches 658; Indels 182; Gaps 51;

QY 92 YDEMFGARSSFFVKGSGVASFSPAGYLTLYREAKDLHFSSAYHLNRRPDLADLTLS 151  
 Db 4 YNNOFSGRASQYVAPGTVSSMFSPPAAYLTLYREARNLHASDSVYLLDTRRPDLKSMALS 63  
 QY 152 QSNMDTEITLTLSNELLEHI-TRKTGSDSALMESLTYRQAIIDTPVHQPYETIROVI 210  
 Db 64 QONMDIELTSLSLNELLESIKTESKLENTTKVEMELSTFRPSGATPYHDAYENVREVI 123  
 QY 211 MTHDSTLSALSRNPVGMQAGSAILAILANISPELYNLTETETEKADALFAQNFSEN 270  
 Db 124 QLQDPGLEQLNASPAIAGLMHQASILLGINASISPELFNLTETETEGNAEELYKKNFG-N 182  
 QY 271 ITPENFASQSWIAKYGLELSEVQKYLGNLQ-N-GYSDSTSAVDNISGLVWNNESKLEA 329  
 Db 183 IEPASLAMPYELKRYNLSDEELSQFICKASNFQGOEYSN---NQLITPVWNSDGTVKV 239  
 QY 330 YKTR-VKTDYDKNINYPDLMEYGNQFFIRANFKVSRFGAT-----LRKN 376  
 Db 240 YRITREYTTNAYQMDVELFP--FGEN--YALDYKFNKFNYNASVLSIKLNDKRELVRTE 294  
 QY 377 AGPSGIVGSLSGPLIANTNPKSNVLSNISDSEYKNGVKIYAYRYTSTTSATNOGGGIFTF 436  
 Db 295 GAQP-----VNIEYSANITLTAD--ISQPFBIGLTRVLPSGWA-YAAAKFTV 340  
 QY 437 ESYPLTTPALKKLNKAIKRLCLTSGLSPNELQTVIRSDNAQGLINDSVLTKVFTLYSHRY 496  
 Db 341 EBYNQYSPLLKLNKAIKRLSRATELSPPTILEGIVRSVNLQDLINTDVLGKVLTKYMQRY 400  
 QY 497 ALSFDDAQVANGSVINQYADDSDSHFNRLFNTPPLKGIPEADGNTVSIIDPDEQSTFA 556  
 Db 401 AIHAETALILCNAPISQRSDYNQPSQFQRLFTPLNLCQYFSTGDEEIDLN-SGSTGDWR 459







Db 1636 KDPHGFSLDDHKTFSGLSQAQKN-----DSEPMBDFSGANALYFWELP 1680  
Qy 1807 YYPMMCFQRLLQSKQDEATQWYNYVNPAGYVINGEIIAPWVNCRPLEBETTSWNP 1866  
Db 1681 YYPMMMAHRLQRFQNFDAANHWPRYVWSPSYVDGKIATYHWNVRPLEEDTSSNAQOL 1740  
Qy 1867 DAIDPDVAQNDPMHYKIATFMRLLDOLILRGDMAYRELTFDALNEAKMVMVTRLELIGD 1926  
Db 1741 DSTDPDAVAQNDPMHYKVATPMATLDDLKMGDAAYQLERDYLAEAKMVMYTQALNLG 1800  
Qy 1927 EPEDYGSQWAAAPSLGSAASQVQAAYQQDLTML 1960  
Db 1801 EPQVWLSTTWANPFLGNAASKTTQOVRAQVLTQL 1834

RESULT 8  
US-08-851-567B-55  
; Sequence 55, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: ffrrench-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Orr, Gregory L.  
; APPLICANT: Roberts, Jean L.  
; APPLICANT: Strickland, James A.  
; APPLICANT: Guo, Lining  
; APPLICANT: Ciche, Todd A.  
; APPLICANT: Sukhapinda, Kitisi  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESS: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851.567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.93804  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 55:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 573 amino acids  
; TYPE: amino acids  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-851-567B-55  
Query Match 12.5%; Score 1632; DB 2; Length 573;  
Best Local Similarity 55.3%; Pred. No. 1.3e-120;  
Matches 317; Conservative 105; Mismatches 147; Indels 4; Gaps 2;  
Qy 1969 LRTANSLVGLPEYNPALTDYMQTLRLRLNLRHNLSDIQPLSLAIYAEPTDPKALLT 2028  
Db 1 LGTANSLTALFLPQENSKLKGWRTLAQRMFNLRHNLSDIQPLSLPLYAKPADPKALLS 60  
Qy 2029 SMYQASQCGSAVLPGTSLSYRPPVWLERTRNLVAQLTQFGTSLLSMAEHDDADLTLL 2088  
Db 61 AAVSASQGGADLPKAPLTIHRFPQMLEGARGLVNQLIQFGSSLLGYSERQDAEAMSQL 120  
Qy 2089 QCGMELATQSIHQRTVDVDAIVLAESRRSAQNRLEKYQOLYDDEINHGQRAMSL 2148  
Db 121 TQASELILTSIRMQDNLAEBSKTKALQVSLAGVQORFDSYSQLYEENINAGQRALAL 180  
Qy 2149 LDMAGQSLAGQVLSIAEGVADLVPNVFGACGSRMGAAALRASAVMSLSATASQYSAD 2208  
Db 181 RSESAIESQGAQISRMAGAGVDMAPNIFGLADGGMHYGATAYATADGIELSASAKWDAE 240  
Qy 2209 KISRSEAYRRRQEWETQRDNAGEVQMDAQLESKIRREAAQOMQVEYQETQQAHTQAO 2268  
Db 241 KVAQSHYRRRQEWKIQRDNAQAEINQLNAQLESLSIRREAAEMQKEYLKTQQAQAAQ 300  
Qy 2269 LELLQRFKTKALYSVMRGLSAIYQFDFLTQSFCLMAQEARRELTNDGVTFIRGAW 2328  
Db 301 LTFURSKFSNQAQYSLWRGLSGIYFQFDLAVSRCLMAQESQYOWEANDNSISFVKFGAW 360  
Qy 2329 NGTTAGLMAGETLLNLNAEMEKVWLERDERALEVTRTVSLAQFYQAL-SSDNFNLTKLT 2387  
Db 361 QGTVAGLLCGEALIQNLAEAEVYLKWSRALEVERTVSLAVVYDSLEGDRFNLAEOIP 420  
Qy 2388 QFLREGKGVGASGNELKSNRQIEASVRLSDLIKIPSDYPESL---GNTRQLKQVSVTL 2444  
Db 421 ALLDKGEGTAGTKENGSLANAILASVKSLSDLKLTGTDYDPSIVGSKNVRRIKQISVSLP 480  
Qy 2445 ALVGPVEDIRAVLNYGGSIVMPRGCSAIALSHGVNDSQGFMLDFNDSRYLPFPEGISVND 2504  
Db 481 ALVGPYQDVQAMLSYGGSTQLPKGCSALAVSHGVNDSQGFQDLPNDGKYLPEFEGIALDD 540  
Qy 2505 GSLLTSPFDATDROKALLESLSLIIHRYTIR 2537  
Db 541 GTLNLQFPNATDKQKAILQTMSDIILHRYTIR 573

RESULT 9  
US-08-851-567B-51  
; Sequence 51, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: ffrrench-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merlo, Donald J.



APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Ciche, Todd A.  
APPLICANT: Sukhapinda, Kitisri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Dow AgroSciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.93804  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 51:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 579 amino acids  
TYPE: amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-51

Query Match 12.4%; Score 1615; DB 2; Length 579;  
Best Local Similarity 54.5%; Pred. No. 3.1e-119;  
Matches 316; Conservative 108; Mismatches 144; Indels 12; Gaps 4;

Qy 1969 LRTANSLVGLPEVNPALTYWTLRLFNRLNLSIDGQPLSLAIYASPTPKALLT 2028  
Db 1 LRSANTLDFLPQINVMYMWQTLQORVNLRLNLSIDGQPLSLAIYATPADPKALLS 60

Qy 2029 SMVQASQGSVALPCTLXLPFPVLMRLNVAQLTQFTSLISMASHDDADELTLLLL 2088  
Db 61 AAVATSGGGKLPESFMSLWFPFMLENARGVSVLTQFTSLQNIERQAEALNALLO 120

Qy 2089 QQGMLATQSIQRTQVDEVDADIAVLAEGRSSAQNRLKYQQLYDEIDNHGQRAMSL 2148  
Db 121 NQAEELILNLSIQDKTEELDAEKTLEKSKAGASRFSYVKLYDENINAGENQAWTL 180

Qy 2149 LDAAGQSLAGVLSIAEGVADLVNPFVGLACGSRWGAALRASASVMSLSATASQYSAD 2208  
Db 181 RASAAGLTAVQASRLAGAAADLVNIFGFGGSRWGAIAEATGYVMFSANVMNTAD 240

Qy 2209 KISRSEAYRRRRQEWIQRDNADGEVKQMDAQLESKLIRREAAQMQVEYQETQQAHTQAO 2268  
Db 241 KISQSEYRRRRQEWIQRDNADGEVKQMDAQLESKLIRREAAQMQVEYQETQQAHTQAO 300

Qy 2269 LELLQKFTNKLALYSWMRGKLSAIYYQFPDITQSFCLMAQEAALRELTQNGVTFIRGGAW 2328  
Db 301 LAFQKQFNSQALYNLWLRGLAAIYFYQYDLAVARCLMAEQAYRWELNDDSAKFIKPGAW 360

Qy 2329 NGTTAGLMAGETLLNLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSSDN--FNLTKL 2386  
Db 361 QGTYAGLLAGETLLMLSLAQMEDAHKRDRALEVERTVSLAEVYAGLFPDNGPFPFLADEI 420

Qy 2387 TQFLREGKGNVGASGNELK-----SNRQIEASVRLSLDKIFSDYPESLGNTRQLKQVSV 2441  
Db 421 DKLVSQGSAGSGNNLAFGAGTDTKTSLQASVSFADLKIREDPYPSLGKIRRIKQISV 480

Qy 2442 TLPALVGYEDIRAVLNYGGSIVMPRGCSATASHGVNDSQFMLDFNDSRYLPPREGISV 2501  
Db 481 TLPALLGYQDVQAILSYGDRAGLANGEALAVSHGMNDSQFQDFDNDFKPLPPEGIAI 540

Qy 2502 NDSGLTSLSPPDAT-----DROKALLESLSDIILHRYTIR 2537  
Db 541 -DQGTLLSFPNASWPEKQKQATMLKTLNDIILHRYTIK 579

RESULT 10  
US-08-851-567B-59  
; Sequence 59, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatis, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: Ifrench-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merio, Donald J.  
; APPLICANT: Orr, Gregory L.  
; APPLICANT: Roberts, Jean L.  
; APPLICANT: Strickland, James A.  
; APPLICANT: Guo, Lining  
; APPLICANT: Ciche, Todd A.  
; APPLICANT: Sukhapinda, Kitisri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Dow AgroSciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255



RESULT 11  
 US-08-851-567B-26  
 ; Sequence 26, Application US/08851567B  
 ; Patent No. 6528484  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ensign, Jerald C  
 ; APPLICANT: Bowen, David J  
 ; APPLICANT: Pettit, James  
 ; APPLICANT: Patig, Raymond  
 ; APPLICANT: Schoonover, Sue  
 ; APPLICANT: French-Constant, Richard  
 ; APPLICANT: Rocheleau, Thomas A.  
 ; APPLICANT: Blackburn, Michael B.  
 ; APPLICANT: Hey, Timothy D.  
 ; APPLICANT: Merlo, Donald J.  
 ; APPLICANT: Orr, Gregory L.  
 ; APPLICANT: Roberts, Jean L.  
 ; APPLICANT: Strickland, James A.  
 ; APPLICANT: Guo, Lining  
 ; APPLICANT: Ciche, Todd A.  
 ; APPLICANT: Sukhapinda, Kitisri  
 ; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
 ; NUMBER OF SEQUENCES: 88  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Dow Agrosciences Patent Department  
 ; STREET: 9330 Zionsville Road  
 ; CITY: Indianapolis  
 ; STATE: IN  
 ; COUNTRY: US  
 ; ZIP: 46268  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/851,567B  
 ; FILING DATE: 05-MAY-1997  
 ; CLASSIFICATION: 514  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/063,615  
 ; FILING DATE: 18-MAY-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/395,497  
 ; FILING DATE: 28-FEB-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/007,255  
 ; FILING DATE: 06-NOV-1995  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/608,423  
 ; FILING DATE: 28-FEB-1996  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/705,484  
 ; FILING DATE: 28-AUG-1996  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.93804  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 608-251-5000  
 ; TELEFAX: 608-251-9166  
 ; INFORMATION FOR SEQ ID NO: 26:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1189 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-851-567B-26

Query Match 11.9%; Score 1553; DB 2; Length 1189;  
 Best Local Similarity 27.0%; Pred. No. 9.9e-114;  
 Matches 447; Conservative 225; Mismatches 477; Indels 504; Gaps 45;

QY	916	QTLAENMEAGLSTQQAOTLADYTAERLSSVLGNWFLANIQPEGVSLHSDRDLYSFLIDN	975
Db	8	QTLKEARDAL-----VAHYIATQVPADLKE-----SIQTADDLYEYLLD	49
QY	976	QVSSAIKTRLAETAGIQLYINRAL-----NRIENARADVSTROFFFTDM-TVNNRYSTW	1030
Db	50	KISDLVTTSPLEAETAGSLQFLHRAIEGYDGTLDASAKPYFADEQFLYNWDSFNHRYSTW	109
QY	1031	GGVSRVLYPENYIDPTQRIQOTRMDELLENISQSLSRDTEVEDAFKTYITFRFTVADL	1090
Db	110	AGKERLKEYAGDYIDPTLRANKTEIFTAFEGISQGLKSELVESKLRDLYLSYDTLATL	169
QY	1091	KVVSAYHDNVSNTGLTWFGVGTRENLEPEYVNRVNDISRMQAGELAAANAKWETKIDTAV	1150
Db	170	DYITACQCKDKNT---IFFIGRTQNPAYFYWRKLTIV-TDGGKLPQWSEWRAINAGI	225
QY	1151	N-PYKDAIRPVI FRERLHLI WKEEVAKNGTDPVETVDRFTLKLAFRLHDSWSAPWSY	1209
Db	226	SEAYSGHVEFPFENNKLHWR-----FTI-----	249
QY	1210	DITTVQEAVTDKKPDTERLALAAAGFQGEDTLIVFYVYTKGKSYSDFGGSKNKNVAGMTI-Y	1268
Db	250	-----SKEDKIDFVY-----KNIWMSSDY	269
QY	1269	GDGFKKMENTALSRYSQLKNTFDIIHTQGNLDVRKASYRFAQDPEVPASLNGSAIGDD	1328
Db	270	SWASKKILELSFTDYNRVGAT-----GSSSPTEVASQYGS-----AQMNISD-----DG	315
QY	1329	SLTWEN--GNIPQITSKYSSDNLAITLHNAAFVRYDGSNVIRNKQISAMKLTGVGDK	1386
Db	316	TVLIFQNAAGATP-----STGVTLCYD-SGNVKN-----L	345
QY	1387	SOYGNAFIANTVKHYGYSGLGGPITVYNKTKNYIASVQGHLMNADYTRRLILTPVENN	1446
Db	346	STGSANLSSKD-----YATTK-LRMCHGQSYN-----DNN	375
QY	1447	YYARLFEPFPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGQFISSVQSSGWLDD	1506
Db	376	Y---CNFTLSINTI-----EFTSY-----	391
QY	1507	TGINNTDIKITVMAGSKTHTTASDHASLPANSFDAMPYTFKPLEIDASSLAFTNNIAP	1566
Db	392	-----GTFSGKQFTPPSG-----	407
QY	1567	LDIVFETKADGRVLGKIKQTLVSKRVNVPEDILFLRETHSGAQYMLGVVYRIRLNTILL	1626
Db	408	-----AIDLHLFNY-----VDLNLALL	423
QY	1627	ASQLVSRANTGIDTILTWETQRLPEPPLGEGFFANFVLPKYDPAEHDGDERWFKIHGNVG	1686
Db	424	-----DISLDSLLNYDVQ-----GQFG	440
QY	1687	GNTGQPYSGMLSDTSETMTLFPVYAEGYMHGVLGVGVQKITYDNTWESAFPFYD	1746
Db	441	G-----	441
QY	1747	ETKQOFLINDADHDSGMTQCGIVKNIKKYKGLNVSIAATGYSAPMD-FNSASALYYWEL	1805
Db	442	-----SNPVDNFGPYGIYLWEI	459
QY	1806	PYYTPMCFQRLLOEKQDEATQWYNYNYPAGY-IVNGEI-----APWNCRPLEETT	1859
Db	460	FFHIPFLVTRMQTEQRYEDADTWYKIFRSAGYRDANGQLIMDGSKPKRYWNVPLQD	519
QY	1860	SWANPLDAIDPDVAQNDPMHYKATFWRLDQLILRGDMAYRELTDALNEAKWYVR	1919
Db	520	AWDTTQPATTPDVVIAMADPMHYKLAIFLHTLDLIIARGDSAYROLERDTLVEAKNYIQ	579
QY	1920	TLELLGDEPDYGSQOWAAPSL-----GAASQTVQAAVQODLTM-----LGRGGVSKNLRTAN	1973
Db	580	AQQLLGPDPDHTTNTWNPFLSKAGAIATFTFLSSPEVMTFAAWLSAGD-----TAN	633

Qy 1974 SLVGLFPEYNPALTDYQWTLRLRLNHLNSIDGQPLSLAIYAEPDTPKALLTSMVQA 2033  
Db 634 IGDGDFLPFYNDVLGYWQKLELRLYNLRLNLSLDGQPLNPLVATPVPDKTL--QRQQA 691  
Qy 2034 SQGSAVLP-----GTLSLYRFPVMLERLNLVAQLTOFGTSLLSMAEHDDADELTLLIQ 2089  
Db 692 GGDGTGSSPAGQSGVQWRYPPLVERARSASVLLTQFNSLQTTLEHQDNKMTILLQT 751  
Qy 2090 QGMELATQSIHQRTVDEVDADIAVLAESRRSAQNLEKYQQLYDEDDINGEQRAMSL 2149  
Db 752 QQEAILKHQHDIQNNLKGSLTALQASRDGDTLQKHYSDLINGLSAAETAGTLR 811  
Qy 2150 DAAGQSLAGQVSLTAEGVADLVNPFVGLCGSRWGAALRASASVMSLSATASQYADK 2209  
Db 812 STAMITGVATGLLIAGGIANAVPNFGLNGSGEWGAPLIGSGOATQVGAGIQDQAGI 871  
Qy 2210 ISRSEATRRRQEWIQRDNADGEVKQMDAQLESILKIRREAAQMQVEYEQTAQHTQAL 2269  
Db 872 SEVTAGYQRRQEWALQRIADNEITQLDAQISLQEQITWAQKQITLSETEQANAQAIY 931  
Qy 2270 ELLQKFTNKALYSWMRGKLSAIYYQFPLDTQSFCLMAQEARRELTLD--NGVTFIRGAW 2328  
Db 932 DLQTTRETGQALYNMAGRLSALYYQMYDSTLPLCLQFKAALVQELGKESDSLFQVPVW 991  
Qy 2329 NGTTAGLMAGTLLNLNAEMKVLDERALEVTRTVLSAQFYQALSSDNFNLTXTLQ 2388  
Db 992 NDLWQGLLAGELSELEQLDAIWLARGGIGLEAIRVSLDTLP-----GTGILSENINK 1046  
Qy 2389 FLREGKNGVAGSNELKLSNNRQIEASVRLSLDKLPSDYPESLGN--TRQLKQSVTLPAL 2446  
Db 1047 VL-NGEYVSPSGVTLTGDIFQATLDSQLGLDQSY--NLGNEKKRRIKRIAVTLPTL 1103  
Qy 2447 VGPYEDTRAVLNYGGSIVMPGCCAIALSHGVNDSGQPMDFNDSRVLPEGISVNDSGS 2506  
Db 1104 LGPYQDLATLVMAEIA-----ALSHGVNDGGRFVTDNDSRFLPPEGRDAT-TGT 1154  
Qy 2507 LTLSPFDA--TDROKALLESLDIILHURYTR 2537  
Db 1155 LELNIFRAGKEGTQHELVANLSDIIVHLNYIR 1187

## RESULT 12

US-08-851-567B-37  
; Sequence 37, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: French-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merio, Donald J.  
; APPLICANT: Orr, Gregory L.  
; APPLICANT: Roberts, Jean L.  
; APPLICANT: Strickland, James A.  
; APPLICANT: Guo, Lining  
; APPLICANT: Ciche, Todd A.  
; APPLICANT: Sukhupinda, Kitisri  
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow AgroSciences Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: US  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/851,567B  
; FILING DATE: 05-MAY-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/063,615  
; FILING DATE: 18-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/395,497  
; FILING DATE: 28-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/007,255  
; FILING DATE: 06-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,423  
; FILING DATE: 28-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/705,484  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.93804  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 37:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 845 amino acids  
; TYPE: amino acids  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein (partial)  
US-08-851-567B-37

Query Match 10.7%; Score 1401.5; DB 2; Length 845;

Best Local Similarity 36.9%; Pred. No. 6e-102;  
Matches 328; Conservative 167; Mismatches 238; Indels 95; Gaps 27;

Qy 566 VNSGELYQLGKLAGVLDQAQNTITLSVEFVSSLYRLTLARVQLTVNELCMYGLSPFNG 625  
Db 4 IDVSLFRLKLTIDHDNKGDKIKNNLKNLSLYIGKLLADHQLTIDELDLLL-IAVGEG 62  
Qy 626 KTTAS-LSSGELPLVILVYQVTQWLTEARITTEAIWLLCTPSPSGNISPEISLNLNLR 684  
Db 63 KTNLSAISDKQLATLIRKLTITISWLTQKWSVQFLFIMTSTSYNKTITPEIKNLLDTVY 122  
Qy 685 PSI---SEDMAQSHNRLEQAEIILAPFTAAATLHLASPDMAVYILWTDNLRP---CGLDIAG 739  
Db 123 HGLQGFQKDKAD-----LHVMAPYIAATLQLSSENVASHVLLWADKLPQGDGAMTAEG 176  
Qy 740 FM-----TLVLKESLNANETTTQ-LVQFCHVMAQLSLSVQTLRLSEALSVLVISGFVLGA 794  
Db 177 FWDMLNTKYPGSGSEAVETQEHIVQYCOALAQLAEMVYHSTGINENAFRLVTKP-EMFGA 235  
Qy 795 KNQPAQOHTDITLPSLYRPFQWINGLGNPGSDTLDMLRQOTLTADRSLASVWGLDSMVQ 854  
Db 236 ATGAAPAHDAALSILMLTRFADWVNALGEKASSVLAFAEANSLTAEQLADANLNLQL 295  
Qy 855 AMVSA-----GVNQLQCQMODINTVLQWIDVASALHTMPSVIRTLVNIY-----VT 900  
Db 296 ASIOAQNHQHLPPVTPENAFSCWTSINTILQWNVVAQLKCRPT-----GRFREGAGLY 350  
Qy 901 ALNKAESNL-PSWDEWQTLAENMEAGLSTQQAOTLADYTA---ERLSVLCNWFANIQP 956  
Db 351 SINERDITLCPVGRKRIRNRVEFN-----NRLIHNAFLDESRSAAALSTYYIRQVAK 404  
Qy 957 EGVSLHSRDDLVSFLIDNQVSSAIKTRLAETAGIQLYINRALNIEPNARADVSTRO 1016  
Db 405 AAAAIIKSRDLDLYQLLLIDNQVSSAIIKTRIAETASITLYVNRALENVEENANSVIRSQ 464



Patent No. 6528484  
GENERAL INFORMATION:  
APPLICANT: Ensign, Jerald C  
APPLICANT: Bowen, David J  
APPLICANT: Petell, James  
APPLICANT: Fatig, Raymond  
APPLICANT: Schoonover, Sue  
APPLICANT: French-Constant, Richard  
APPLICANT: Rocheleau, Thomas A.  
APPLICANT: Blackburn, Michael B.  
APPLICANT: Hey, Timothy D.  
APPLICANT: Merlo, Donald J.  
APPLICANT: Orr, Gregory L.  
APPLICANT: Roberts, Jean L.  
APPLICANT: Strickland, James A.  
APPLICANT: Guo, Lining  
APPLICANT: Ciche, Todd A.  
APPLICANT: Sukhapinda, Kitisiri  
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
NUMBER OF SEQUENCES: 88  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dow Agrosciences Patent Department  
STREET: 9330 Zionsville Road  
CITY: Indianapolis  
STATE: IN  
COUNTRY: US  
ZIP: 46268  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/851,567B  
FILING DATE: 05-MAY-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/063,615  
FILING DATE: 18-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/395,497  
FILING DATE: 28-FEB-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/007,255  
FILING DATE: 06-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,423  
FILING DATE: 28-FEB-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/705,484  
FILING DATE: 28-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.93804  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 627 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-851-567B-28  
Query Match 4.5%; Score 593; DB 2; Length 627;  
Best Local Similarity 20.3%; Pred. No. 7.7e-38;  
Matches 212; Conservative 127; Mismatches 243; Indels 464; Gaps 32;  
916 OTLAENMEAGLSTQAOQLADYABRLSSVLCNWFANIQPEGVSLHRRDLISYFLIDN 975  
8 QTLKEARDAL-----VAHYIAQVPADLKE-----SIQTADDLYEYLLDIT 49

QY 976 QVSSAIKTRTLLAEAIAGIQLYINRAL---NRIEPNARADVSTROFFTDW-TVNNRYSTW 1030  
D 50 KISDLVTISPLSEAIAGSLQFIHRAIEGYDGLTADSAKPYFADEQFLYNNWDSFNHRISTW 109  
QY 1031 GGVSRLLVYPENYIDPTQIRIGOTRMDDELLENISQSKLSRDTVEDAFKTYLTRFETVADL 1090  
D 110 AKERLAFYAGDYIDPTLRNKTEIFTAFBQGISQSKLSSELVESKLRDYLISYDTLATL 169  
QY 1091 KVSAYHDNVSNTGLTWFGOTRENLPFYWRNVDISRMQAGELAAANAKWTKIDTAV 1150  
D 170 DYITACQCKDNKT---IFFIGRTQNAFYAFYWRKLTIV-TDGGKLPDQMSERAINAGI 225  
QY 1151 N-PYKDAIRPVI PRERLHLIWEKEEVAKNCTDPEVYDFRTLLKLAFLRHGDSWSAPWSY 1209  
D 226 SEAYSGHVEFWENKLIIRW-----FTI----- 249  
QY 1210 DITTOVEAVTDKKPDTERLALAAASFGQEDTLLVYVYTKGYSDFGSGSNKNVAGMTI-Y 1268  
D 250 -----SKEDKIDFVY-----KNIVWMSDY 269  
QY 1269 GDGSKFKNMENTALSRYSQLKNFTDIHTQGNLIVKASYRPAQDFEVPASLNMGSAIGDD 1328  
D 270 SWASKKILELSFTDYNRVGAT-----GSSSPTEVASQYGS-----AQMNISD---DG 315  
QY 1329 SLTWEN--GNIPOITSKYSSDNLAITLHNAAFVRYDGSNGVIRNKKQISAMKLTGVGDK 1386  
D 316 TVLIFQAGGATP-----STGVTLCYD-SGNVIKN-----L 345  
QY 1387 SOYGNAFIIANTVKHYGYSDLGGPITVYANKTNYIASVOGHLNADYTRRLIITPVENN 1446  
D 346 SSTGSANLSSKO-----YATTK--LRMCHGQSYN-----DNN 375  
QY 1447 YYARLFFPPSPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFQIFSSYQSSGWLDD 1506  
D 376 Y---CNFTLSINTI-----BFTSY----- 391  
QY 1507 TGINNTDIKITVMAGSKTHFTASDHIASLPANSFDAMPYTFKPLEIDASSLAFTNNIAP 1566  
D 392 -----GTFSSDGKQFTPPSGS----- 407  
QY 1567 LDIVFETKAKDGRVLGKIQTLSVKRVNYPEDILFLRETHSGAQYQMLGVYRIRLNTLL 1626  
D 408 -----AIDLHLFNY-----VDLNAALL 423  
QY 1627 ASQLVSRANTGIDTILTWETQRLPEPLGEGFFANFVLPKYDPAEHGDERWFKIHGNVG 1686  
D 424 -----DISLDSLNYDVQ-----GQFG 440  
QY 1687 GNTGRQPYSGMLSDTSETSMTLFVPYABGYMHGVRGLGVGYQKITDYDNTWESAFFYD 1746  
D 441 G----- 441  
QY 1747 ETQKFVLLINDADHDGSMGTQGIKNIKKYKGFNLVSIATGYSAPMD-FNSASALYYWEL 1805  
D 442 -----SNPVDNFGPYGIYLWEI 459  
QY 1806 FYTTPMCFORLLQKOFDEATQINIVYNPAGY-IVNGEI-----APWIKNCRPLESTT 1859  
D 460 FFHIPFLVTVMQTEORYEDATWYKIFRSAGYRDANGQLIMDGSKRYVNVMPLODIT 519  
QY 1860 SWNANPLDAIDPDVAQNDPMHYKIATFMRLDQLILRGDMAYRELTLDALNEAKWYVR 1919  
D 520 AWDTTQPATDTPDVIMADPMHYKIAIPLHTLDDLLIARGDSAYQLERDITLVEAKWYIQ 579  
QY 1920 TLELIGDEPEDYGSQOWAAPSLSGAA 1945  
D 580 AQLLGLPRPDIIHTTNPENPLSKEA 605  
RESULT 15  
US-08-851-567B-34  
; Sequence 34, Application US/08851567B



Qy 907 SNLPSWDEWQTLAENMEAGLSTQQ 930  
Db 1078 ALLAS-----VAQNLSAATSNRQ 1095

Search completed: February 16, 2006, 21:46:27  
Job time : 69.1714 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 230.217 Seconds  
(without alignments)  
4606.314 Million cell updates/sec

Title: US-10-754-115-34

Perfect score: 13043

Sequence: 1 MYSTAVLNKISPTRDGQTM.....KALLESLDIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA Main:

- 1: /cgm2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgm2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 3: /cgm2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 4: /cgm2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 5: /cgm2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 6: /cgm2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13043	100.0	2538	US-10-609-113-49	Sequence 49, Appl
2	13043	100.0	2538	US-10-753-901-20	Sequence 20, Appl
3	13043	100.0	2538	US-10-754-115-20	Sequence 20, Appl
4	13043	100.0	2538	US-10-754-115-34	Sequence 34, Appl
5	5496.5	42.1	2523	US-10-753-901-14	Sequence 14, Appl
6	5496.5	42.1	2523	US-10-754-115-14	Sequence 14, Appl
7	5043	38.7	2516	US-09-817-514A-2	Sequence 2, Appl
8	5043	38.7	2516	US-10-262-794A-47	Sequence 47, Appl
9	5043	38.7	2516	US-10-706-424-2	Sequence 2, Appl
10	5043	38.7	2516	US-10-703-280-4	Sequence 4, Appl
11	4986.5	38.2	2504	US-09-817-514A-8	Sequence 8, Appl
12	4986.5	38.2	2504	US-10-262-794A-12	Sequence 12, Appl
13	4986.5	38.2	2504	US-10-754-115-59	Sequence 59, Appl
14	4376	38.2	2534	US-11-020-848-11	Sequence 11, Appl
15	4802	36.8	2381	US-10-706-424-8	Sequence 8, Appl
16	4802	36.8	2381	US-10-754-115-63	Sequence 63, Appl
17	4574	35.1	2499	US-10-706-424-4	Sequence 4, Appl
18	4572	35.1	2499	US-10-754-115-62	Sequence 62, Appl
19	4433	34.0	2177	US-10-706-424-6	Sequence 6, Appl
20	3303	25.3	1849	US-10-262-794A-49	Sequence 49, Appl
21	3278.5	25.1	1844	US-10-262-794A-53	Sequence 53, Appl
22	2655	20.4	1273	US-11-020-848-6	Sequence 6, Appl
23	1643	12.6	1205	US-10-609-113-5	Sequence 5, Appl
24	1632	12.5	573	US-10-262-794A-55	Sequence 55, Appl
25	1615	12.4	579	US-10-262-794A-51	Sequence 51, Appl
26	1591	12.2	1565	US-10-262-794A-59	Sequence 59, Appl
27	1572.5	12.1	1208	US-10-609-113-9	Sequence 9, Appl

## ALIGNMENTS

### RESULT 1

```

US-10-609-113-49
; Sequence 49, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; TITLE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 2538
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-609-113-49

```

```

Query Match      100.0%; Score 13043; DB 4; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSTAVLNKISPTRDGQTMADLQYLSFSELRLKIFDDQLSWGAEARHLYHETIQQKN 60
DB 1 MYSTAVLNKISPTRDGQTMADLQYLSFSELRLKIFDDQLSWGAEARHLYHETIQQKN 60

QY 61 RLLEARIIFRANPOLSGAIRLGIERSVRSYDENMFGARSSSFVKPGSVASMFSPAGYLT 120
DB 61 RLLEARIIFRANPOLSGAIRLGIERSVRSYDENMFGARSSSFVKPGSVASMFSPAGYLT 120

QY 121 ELYREAKDLHFSSSAHYLDNRPPDLADLTLSQSNMDTETISTLTLSNELLLEHITRTGDD 180
DB 121 ELYREAKDLHFSSSAHYLDNRPPDLADLTLSQSNMDTETISTLTLSNELLLEHITRTGDD 180

QY 181 SDALMESLSTYRQAITDTPYHQPYETIRQIVIMTHDSTLSALSNPEVMQAGASLALIA 240
DB 181 SDALMESLSTYRQAITDTPYHQPYETIRQIVIMTHDSTLSALSNPEVMQAGASLALIA 240

QY 241 NISPELYNLTETETKADALFAQNFENITPENFASQSWIAKYGYGSELSEVQYKLGML 300
DB 241 NISPELYNLTETETKADALFAQNFENITPENFASQSWIAKYGYGSELSEVQYKLGML 300

```

QY 301 QNGYSDTSAYVDNI STGLVNNESKLEAYKI TRVKTDYDKNINYFDLMYEGNNOFFIR 360  
DB 301 QNGYSDTSAYVDNI STGLVNNESKLEAYKI TRVKTDYDKNINYFDLMYEGNNOFFIR 360  
QY 361 ANFKVSRFEGATLRKNAGPSIGVLSGPLIANTNFKSNYLSNDSSEYKNGVKIYAYRY 420  
DB 361 ANFKVSRFEGATLRKNAGPSIGVLSGPLIANTNFKSNYLSNDSSEYKNGVKIYAYRY 420  
QY 421 TSSTSATNQGGI FTTFESYPLTITIPALKNKAIRCLTSGLSPELQITIVRSDNAQGIIND 480  
DB 421 TSSTSATNQGGI FTTFESYPLTITIPALKNKAIRCLTSGLSPELQITIVRSDNAQGIIND 480  
QY 481 SVLTKVFYTLFPYSHRYALSFDQAOLVNGSVINQVADDDSVSHFNRLNTPPLKGIPEAD 540  
DB 481 SVLTKVFYTLFPYSHRYALSFDQAOLVNGSVINQVADDDSVSHFNRLNTPPLKGIPEAD 540  
QY 541 GNTVSIIDPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600  
DB 541 GNTVSIIDPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600  
QY 601 TLLARVHQLTVNELCMLYGLSPFNKKTASLSGSELPRPLVILWYQVOWLTTEABITTEAI 660  
DB 601 TLLARVHQLTVNELCMLYGLSPFNKKTASLSGSELPRPLVILWYQVOWLTTEABITTEAI 660  
QY 661 WLLCTPFSGNISPEISNLLNLRPSISEDMAQSHNRELOAEIILAPFTIATLHLASPDMA 720  
DB 661 WLLCTPFSGNISPEISNLLNLRPSISEDMAQSHNRELOAEIILAPFTIATLHLASPDMA 720  
QY 721 RYILLWTDNLRPGGLDLAGFMTLVLBESLNANETTLQVQFCHVMAQSLSVQTLRLSEAE 780  
DB 721 RYILLWTDNLRPGGLDLAGFMTLVLBESLNANETTLQVQFCHVMAQSLSVQTLRLSEAE 780  
QY 781 LSVLVI SGFVLGAKNQOPAGQHNIDTLPFLYRFHQWINGLNGPQSDTLDMLRQOQTLADR 840  
DB 781 LSVLVI SGFVLGAKNQOPAGQHNIDTLPFLYRFHQWINGLNGPQSDTLDMLRQOQTLADR 840  
QY 841 LASVWGLDLSMVTQAMYSAGVNOJCQWODINTVLQWIDVASALHTMPSVIRTLVNI RVT 900  
DB 841 LASVWGLDLSMVTQAMYSAGVNOJCQWODINTVLQWIDVASALHTMPSVIRTLVNI RVT 900  
QY 901 ALNKAESNLPDSEWQTLAENMEAGLSLQQAQTLADYTAERLSVLCNWFLANIQPEGV 960  
DB 901 ALNKAESNLPDSEWQTLAENMEAGLSLQQAQTLADYTAERLSVLCNWFLANIQPEGV 960  
QY 961 LHSRDDLYSYFLIDNQVSSAIKTRLABAIAIGQLYINRALNRIEPNARADVSTRQFPTD 1020  
DB 961 LHSRDDLYSYFLIDNQVSSAIKTRLABAIAIGQLYINRALNRIEPNARADVSTRQFPTD 1020  
QY 1021 WTVNRRYSTWGVSRVLVYPENYIDPTQRIQOTRMDBELLENISQSKLSRDTVEDAPKTY 1080  
DB 1021 WTVNRRYSTWGVSRVLVYPENYIDPTQRIQOTRMDBELLENISQSKLSRDTVEDAPKTY 1080  
QY 1081 LTRFETVADLKVSAYHDNVNSNTGLTWFGVQCTRENLPYYWRNVDISRMOAGELAANAW 1140  
DB 1081 LTRFETVADLKVSAYHDNVNSNTGLTWFGVQCTRENLPYYWRNVDISRMOAGELAANAW 1140  
QY 1141 KEWKIDTAVNPYKDAIRPVI FRBRHLHIWVEKEBAVAKNGTDPVETTYDRFTLKLAFLRHD 1200  
DB 1141 KEWKIDTAVNPYKDAIRPVI FRBRHLHIWVEKEBAVAKNGTDPVETTYDRFTLKLAFLRHD 1200  
QY 1201 GSWAPASYSIDLTTOVEAVTDKPPDTERLALAASFGQGEDTLVLVFPYKTKGSYDFGGSNK 1260  
DB 1201 GSWAPASYSIDLTTOVEAVTDKPPDTERLALAASFGQGEDTLVLVFPYKTKGSYDFGGSNK 1260  
QY 1261 NVAGMTIYGDGSPFKMENTALSRYSQLKNPTFDIHTIQGNDLVRKASRYFAQDFEVPASLN 1320  
DB 1261 NVAGMTIYGDGSPFKMENTALSRYSQLKNPTFDIHTIQGNDLVRKASRYFAQDFEVPASLN 1320  
QY 1321 MGSAGDDSLTVMENGNIPQITTSKYSSDNLAITLHNAAFVRYDGSNVIRNKQISAMKL 1380  
DB 1321 MGSAGDDSLTVMENGNIPQITTSKYSSDNLAITLHNAAFVRYDGSNVIRNKQISAMKL 1380  
QY 1381 TGVDGKSYGNAFIANTVVKHYGSDLGGPITVYNKTKNIVASVOGHLMNADYTRRLIL 1440

DB 1381 TGVDGKSYGNAFIANTVVKHYGSDLGGPITVYNKTKNIVASVOGHLMNADYTRRLIL 1440  
QY 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGNKTSDFKFKCSYAVDGNNSQGFQIFSSYQSS 1500  
DB 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGNKTSDFKFKCSYAVDGNNSQGFQIFSSYQSS 1500  
QY 1501 GMLDIDGGINNTDIIKITVMAGSKTHHTFTASDHTASLPANSFDAMPYTFKPLEIDASSIAP 1560  
DB 1501 GMLDIDGGINNTDIIKITVMAGSKTHHTFTASDHTASLPANSFDAMPYTFKPLEIDASSIAP 1560  
QY 1561 TNNIAPLDIIVFETKAKDGRVLGKIKOTLSVKRVNYPNEDILFLRTHESGAQYMLGVYRI 1620  
DB 1561 TNNIAPLDIIVFETKAKDGRVLGKIKOTLSVKRVNYPNEDILFLRTHESGAQYMLGVYRI 1620  
QY 1621 RLNTLLASQLVSRANTGIDITILNMQTORLPEPPLEGGEFFANFVLPKYDPPAEHGBERWPKI 1680  
DB 1621 RLNTLLASQLVSRANTGIDITILNMQTORLPEPPLEGGEFFANFVLPKYDPPAEHGBERWPKI 1680  
QY 1681 HIGNVGNTGROPYSGMLSDTSETSMTLFPVYAEGYMHGEGVRLGVGYQKITYDNTWES 1740  
DB 1681 HIGNVGNTGROPYSGMLSDTSETSMTLFPVYAEGYMHGEGVRLGVGYQKITYDNTWES 1740  
QY 1741 AFFYFDETKQOFVLINDADHDHDSGMTQOGIVKNIKKYKGFNLVSIATGYSAPMDFNSASAL 1800  
DB 1741 AFFYFDETKQOFVLINDADHDHDSGMTQOGIVKNIKKYKGFNLVSIATGYSAPMDFNSASAL 1800  
QY 1801 YYWELFYTTMPCFORLQEKQFDEATOWINVYNPAGIYVNGBIAPWINCRLPLEETTS 1860  
DB 1801 YYWELFYTTMPCFORLQEKQFDEATOWINVYNPAGIYVNGBIAPWINCRLPLEETTS 1860  
QY 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920  
DB 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920  
QY 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGVSKNLTANSILVGLFL 1980  
DB 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGVSKNLTANSILVGLFL 1980  
QY 1981 PEYNPALTDYQWTLRLRFLNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGGSAV 2040  
DB 1981 PEYNPALTDYQWTLRLRFLNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGGSAV 2040  
QY 2041 LPGTSLYRFPVMLERTNRLVAQLTQFTGTSLSMAEHDDADELTLLLOQGMELATOSIR 2100  
DB 2041 LPGTSLYRFPVMLERTNRLVAQLTQFTGTSLSMAEHDDADELTLLLOQGMELATOSIR 2100  
QY 2101 IQQRTVDEVDADIIVLAESRRSAQNRLKEYQQLYDEIDINHGEQRAMSLDDAAAGQSLAQ 2160  
DB 2101 IQQRTVDEVDADIIVLAESRRSAQNRLKEYQQLYDEIDINHGEQRAMSLDDAAAGQSLAQ 2160  
QY 2161 VLSIARGVADLVPNVFGIACGSRWGAALRASAVMSLSATASQYADKISRSEAYRRRR 2220  
DB 2161 VLSIARGVADLVPNVFGIACGSRWGAALRASAVMSLSATASQYADKISRSEAYRRRR 2220  
QY 2221 QEWIQRDNADGEVKQMDAQLESUKIRREAAQMOVYQETQQAHTQOALELQKFTNKA 2280  
DB 2221 QEWIQRDNADGEVKQMDAQLESUKIRREAAQMOVYQETQQAHTQOALELQKFTNKA 2280  
QY 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEALRRELTDNGVTFFIRGGAWNGTTAGIMAGET 2340  
DB 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEALRRELTDNGVTFFIRGGAWNGTTAGIMAGET 2340  
QY 2341 LLLNLAEMEKVWLERDRERALEVTRTVSLAQYQALSSDNFNLTBKLTOFLREGKGNVGAS 2400  
DB 2341 LLLNLAEMEKVWLERDRERALEVTRTVSLAQYQALSSDNFNLTBKLTOFLREGKGNVGAS 2400  
QY 2401 GNEUKLSNRQIEASVRLSDLKI FSDYPESLGNTRQLQKQSVTLTPALVGPYEDIRAVLNYG 2460  
DB 2401 GNEUKLSNRQIEASVRLSDLKI FSDYPESLGNTRQLQKQSVTLTPALVGPYEDIRAVLNYG 2460  
QY 2461 GSIYMPGCSAIALSHGVNDGSGQFMPLDFNDSRYLPFEGISVNDSGSLTSLSPFDATDROKA 2520

Db 2461 GSIVMPGCSAIALSHGVNDSQFMLDFNDSRYLPFEGISVNDSSGLTSLSPDATDROKA 2520

Qy 2521 LLESLSDIILHIRYTI RS 2538

Db 2521 LLESLSDIILHIRYTI RS 2538

RESULT 2

US-10-753-901-20

; Sequence 20, Application US/10753901

; Publication No. US200401941641

; GENERAL INFORMATION:

; APPLICANT: Bintrim, Scott

; APPLICANT: Mitchell, Jon

; APPLICANT: Martinua, Ignacio

; APPLICANT: Apel-Birkhold, Patricia

; APPLICANT: Schafer, Barry

; APPLICANT: Bevan, Scott

; APPLICANT: Young, Scott

; APPLICANT: Guo, Lining

; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control

; FILE REFERENCE: DAS-105X

; CURRENT APPLICATION NUMBER: US/10/753,901

; CURRENT FILING DATE: 2004-01-07

; PRIOR APPLICATION NUMBER: US 60/441,717

; PRIOR FILING DATE: 2003-01-21

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 20

; LENGTH: 2538

; TYPE: PRT

; ORGANISM: Xenorhabdus nematophilus

US-10-753-901-20

Query Match 100.0%; Score 13043; DB 4; Length 2538;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYSTAVLNKISPTRDGQTMTLADLQYLSFSELKIFDDQLSWGAEARHLXHTHIOKKN 60

Db 1 MYSTAVLNKISPTRDGQTMTLADLQYLSFSELKIFDDQLSWGAEARHLXHTHIOKKN 60

Qy 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120

Db 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120

Qy 121 ELYREAKDLHFSSAYHLNDRRPDLADLTLSQSNMDETSITLSNELLELHITRKTCGD 180

Db 121 ELYREAKDLHFSSAYHLNDRRPDLADLTLSQSNMDETSITLSNELLELHITRKTCGD 180

Qy 181 SDALMESLSTYRQADTPYHOPYETIROVIMTHDSTLSALSRNPVWMOAGASLLAILA 240

Db 181 SDALMESLSTYRQADTPYHOPYETIROVIMTHDSTLSALSRNPVWMOAGASLLAILA 240

Qy 241 NISPELYNLTETETKKNADALFAQNFSENITPENFASQSWIAKYIGLESEVQKYLQWL 300

Db 241 NISPELYNLTETETKKNADALFAQNFSENITPENFASQSWIAKYIGLESEVQKYLQWL 300

Qy 301 QNGVSDSTSAVDNISTGLVNNESSKLEAYKI TRVKTDDYDKNINYPDLMEYEGNQFFIR 360

Db 301 QNGVSDSTSAVDNISTGLVNNESSKLEAYKI TRVKTDDYDKNINYPDLMEYEGNQFFIR 360

Qy 361 ANFKVSRFEGATLRKNAGPSGLVSGPLANTNFKSNLYLSNIDSSEYKNGVKIYAYRY 420

Db 361 ANFKVSRFEGATLRKNAGPSGLVSGPLANTNFKSNLYLSNIDSSEYKNGVKIYAYRY 420

Qy 421 TSSTSATNQGGGIFTFESYPLTIFALKNKAIRCLTSLGSPNELQTIIVRSNDAQGIIND 480

Db 421 TSSTSATNQGGGIFTFESYPLTIFALKNKAIRCLTSLGSPNELQTIIVRSNDAQGIIND 480

Qy 481 SVLTQVFTLTFYSHRYALSFDDAQVNGSVINQYADDSVSHFNRLFNTPLKGIKIFPAD 540

Db 481 SVLTQVFTLTFYSHRYALSFDDAQVNGSVINQYADDSVSHFNRLFNTPLKGIKIFPAD 540

Qy 541 GNTVSIIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600

Db 541 GNTVSIIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600

Qy 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASISSGELPRLVWLYQVTLWLTAEATITTEAI 660

Db 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASISSGELPRLVWLYQVTLWLTAEATITTEAI 660

Qy 661 WLLCTPFESGNISPEISNLNLRPSISSEDMAQSHNRELQAEILAPFTAATLHLASPDWA 720

Db 661 WLLCTPFESGNISPEISNLNLRPSISSEDMAQSHNRELQAEILAPFTAATLHLASPDWA 720

Qy 721 RYILLWTDNLRPGLDIAAGFMTLVKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780

Db 721 RYILLWTDNLRPGLDIAAGFMTLVKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780

Qy 781 LSVLVISGFVILGAKNQAGOHNDITLFSLRFHQWINGLGNPGSDTTLMLRQOQLTADR 840

Db 781 LSVLVISGFVILGAKNQAGOHNDITLFSLRFHQWINGLGNPGSDTTLMLRQOQLTADR 840

Qy 841 LASVWGLDISMVTQAMVSAGVNLQOCQWODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900

Db 841 LASVWGLDISMVTQAMVSAGVNLQOCQWODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900

Qy 901 ALNKAESNLPSEWDEWQTLAENWEAGLSTQQAQTLADYTAERLSSVLCNWFLANIQPEGVS 960

Db 901 ALNKAESNLPSEWDEWQTLAENWEAGLSTQQAQTLADYTAERLSSVLCNWFLANIQPEGVS 960

Qy 961 LHSRDDLYSYFLIDNQVSSAIKTRLABAIAIGIQLYINRNLNRIEPNARADVSTQFFTD 1020

Db 961 LHSRDDLYSYFLIDNQVSSAIKTRLABAIAIGIQLYINRNLNRIEPNARADVSTQFFTD 1020

Qy 1021 WTVNRRSTWGVSRSLVYPENYIDPTORIGOTRMDDELLENISOSKLSRDTVEADFPTY 1080

Db 1021 WTVNRRSTWGVSRSLVYPENYIDPTORIGOTRMDDELLENISOSKLSRDTVEADFPTY 1080

Qy 1081 LTRFETVADLVKVSAYHDNVNSNTGLTWFGQTRLENLPEYVYWRNVYDISRMOAGELAANAW 1140

Db 1081 LTRFETVADLVKVSAYHDNVNSNTGLTWFGQTRLENLPEYVYWRNVYDISRMOAGELAANAW 1140

Qy 1141 KEWKIDITAVNYPYKDAIRPVI FRERLHLI WYKEEVAKNGTDPVETYDRFTLKLAFLRH 1200

Db 1141 KEWKIDITAVNYPYKDAIRPVI FRERLHLI WYKEEVAKNGTDPVETYDRFTLKLAFLRH 1200

Qy 1201 GSWAPWSYDITTOVEAVTDKPKPTERLALAASFGQEDTLLVFYKTKGYSDFGSGNK 1260

Db 1201 GSWAPWSYDITTOVEAVTDKPKPTERLALAASFGQEDTLLVFYKTKGYSDFGSGNK 1260

Qy 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIIHTQGNDLVRKASYPFAQDFEVPASLN 1320

Db 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIIHTQGNDLVRKASYPFAQDFEVPASLN 1320

Qy 1321 MGSAGDLSLTMENGINI POITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380

Db 1321 MGSAGDLSLTMENGINI POITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380

Qy 1381 TGVGKSYGNAFIANTVKHYGGYSDLGPIITVYKTKNYIASVQGHLMADYTRRLIL 1440

Db 1381 TGVGKSYGNAFIANTVKHYGGYSDLGPIITVYKTKNYIASVQGHLMADYTRRLIL 1440

Qy 1441 TPVENNYIARLFEPPFPFSPENTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFQIFSSYQSS 1500

Db 1441 TPVENNYIARLFEPPFPFSPENTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFQIFSSYQSS 1500

Qy 1501 GWLDDITGINNTDITKITVMAGSKTHFTTASDHIAISLPANSFDMPTTKPLEIDASSLAF 1560

Db 1501 GWLDDITGINNTDITKITVMAGSKTHFTTASDHIAISLPANSFDMPTTKPLEIDASSLAF 1560

Qy 1561 TNNIAPLDIVFETAKOGRVLGKIQTLSVKRVNVPEDILFLRTHSGAOVMQGVTRI 1620

Db 1561 TNNIAPLDIVFETAKOGRVLGKIQTLSVKRVNVPEDILFLRTHSGAOVMQGVTRI 1620

```
Qy 1621 RLNTLLASQVSRANTGIDTILTMETORLPBPPLGEGFPANFVLPKYPDPAEGHDERWFKI 1680
Db 1621 RLNTLLASQVSRANTGIDTILTMETORLPBPPLGEGFPANFVLPKYPDPAEGHDERWFKI 1680
Qy 1681 HIGNVGGNTGRQPYSGMLSDTSETMTLPVPAEGYMHGVRGLGVGYKITDYNTWES 1740
Db 1681 HIGNVGGNTGRQPYSGMLSDTSETMTLPVPAEGYMHGVRGLGVGYKITDYNTWES 1740
Qy 1741 AFFYFDETKQOVLINDADHDGSMGTQOGIVKNIKKYKGFNLVSIATGYSAPMDNFNSAL 1800
Db 1741 AFFYFDETKQOVLINDADHDGSMGTQOGIVKNIKKYKGFNLVSIATGYSAPMDNFNSAL 1800
Qy 1801 YYWELFYTPMWCORLLOEQFOBEATQWIVNVPAGYI VNGHIAPIWNCRPLEETTS 1860
Db 1801 YYWELFYTPMWCORLLOEQFOBEATQWIVNVPAGYI VNGHIAPIWNCRPLEETTS 1860
Qy 1861 WNANPLDAIDPDAVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNANPLDAIDPDAVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920
Qy 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVQAAQODLTMLGRGGVSKNLRATANSLVGLFL 1980
Db 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVQAAQODLTMLGRGGVSKNLRATANSLVGLFL 1980
Qy 1981 PEYNPALTDYQWTLRLRFLNLRHNLSDGQPLSLAI VAEPDTPKALLTSMVQASQGGSAV 2040
Db 1981 PEYNPALTDYQWTLRLRFLNLRHNLSDGQPLSLAI VAEPDTPKALLTSMVQASQGGSAV 2040
Qy 2041 LPGTSLSYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLQQGMELATQSIR 2100
Db 2041 LPGTSLSYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLQQGMELATQSIR 2100
Qy 2101 IQORTVDEVDADIIVLAESRRAQNRLEKYQOQYDEIDINHGQRAQMSLLDAAQCSLAGQ 2160
Db 2101 IQORTVDEVDADIIVLAESRRAQNRLEKYQOQYDEIDINHGQRAQMSLLDAAQCSLAGQ 2160
Qy 2161 VLSIAEGVADIVPNVFGACGSRWGAALRASAVMSLSATASQVSADKISRSAYRRRR 2220
Db 2161 VLSIAEGVADIVPNVFGACGSRWGAALRASAVMSLSATASQVSADKISRSAYRRRR 2220
Qy 2221 QEWEIQRDNADGEVKQMDAQLESKIRREAAQMVQEQETQQAHTQAQLELLQRKFTNKA 2280
Db 2221 QEWEIQRDNADGEVKQMDAQLESKIRREAAQMVQEQETQQAHTQAQLELLQRKFTNKA 2280
Qy 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARRELTNGVTFIRGGAWNGTTAGIMAGET 2340
Db 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARRELTNGVTFIRGGAWNGTTAGIMAGET 2340
Qy 2341 LLLNLAEKVKWLERDERALEVTRTVSLAQFYQALSSDNFNLTEKLTOFLREGKNGV GAS 2400
Db 2341 LLLNLAEKVKWLERDERALEVTRTVSLAQFYQALSSDNFNLTEKLTOFLREGKNGV GAS 2400
Qy 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLTPALVGPVEDIRAVLNYG 2460
Db 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLTPALVGPVEDIRAVLNYG 2460
Qy 2461 GSIWMPRGCSAIALSHGWNDSQGMFLDPNDSRYLPFPGISVNDSGSLTSPDPDATDRKA 2520
Db 2461 GSIWMPRGCSAIALSHGWNDSQGMFLDPNDSRYLPFPGISVNDSGSLTSPDPDATDRKA 2520
Qy 2521 LLESLSDIILHIRTIRS 2538
Db 2521 LLESLSDIILHIRTIRS 2538
```

## RESULT 3

US-10-754-115-20

; Sequence 20, Application US/10754115

; Publication No. US20040208907A1

; GENERAL INFORMATION:

; APPLICANT: Hey, Timothy

; APPLICANT: Schleper, Amanda

; APPLICANT: Bevan, Scott

```
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, BaoLong
; APPLICANT: Mezlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10754,115
; PRIOR FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 20
; LENGTH: 2538
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-754-115-20
```

Query Match 100.0%; Score 13043; DB 4; Length 2538;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 MYSTAVLANKISPRDQGTMTLADLOVLSRSELKIFDDQLSWGEARHLVHETIEQKKN 60
Db 1 MYSTAVLANKISPRDQGTMTLADLOVLSRSELKIFDDQLSWGEARHLVHETIEQKKN 60
Qy 61 RLLEARIPTFRANPOLSGAIRLGIERSVRSYDMDFGARSSSFVKPGSVASMFSPAGYLT 120
Db 61 RLLEARIPTFRANPOLSGAIRLGIERSVRSYDMDFGARSSSFVKPGSVASMFSPAGYLT 120
Qy 121 ELYREAKDLHFFSSAYHLNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Db 121 ELYREAKDLHFFSSAYHLNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Qy 181 SDALMESLSTYRQAIDTPYHQPYETIRQVIMTHDSTLSALSRNPEVWQAGASLLAILA 240
Db 181 SDALMESLSTYRQAIDTPYHQPYETIRQVIMTHDSTLSALSRNPEVWQAGASLLAILA 240
Qy 241 NISPELNYLITEETEKNDALFAQNFSENIITPENFASQSWIAKYVGLSEVQKYLGM 300
Db 241 NISPELNYLITEETEKNDALFAQNFSENIITPENFASQSWIAKYVGLSEVQKYLGM 300
Qy 301 QNGVSDSTSAVVDNISGLVWNESKLEAYKITRKTDDYDKNINFDLMYEGNQFPFIR 360
Db 301 QNGVSDSTSAVVDNISGLVWNESKLEAYKITRKTDDYDKNINFDLMYEGNQFPFIR 360
Qy 361 ANFKVSRFEGNTLRKNAGPSGIVGSLGPIANTNFKSNVLSNISDSEYKNGVKIYARY 420
Db 361 ANFKVSRFEGNTLRKNAGPSGIVGSLGPIANTNFKSNVLSNISDSEYKNGVKIYARY 420
Qy 421 TSSTSATNQGGIIFTFESYPLTIPALKLNKAIRCLTSLGSLSPNELQITVRSDNAQGIIND 480
Db 421 TSSTSATNQGGIIFTFESYPLTIPALKLNKAIRCLTSLGSLSPNELQITVRSDNAQGIIND 480
Qy 481 SVLTQVFTYTLFYSHRYALSFDDAQVLNGSVINQVADDDSVSHFNRLFTPTPLKGIPEAD 540
Db 481 SVLTQVFTYTLFYSHRYALSFDDAQVLNGSVINQVADDDSVSHFNRLFTPTPLKGIPEAD 540
Qy 541 GNTVSDIDPDEEQSTFARSALMRGLGVNSGELYQGLAGVLDAQNTTILSVFVSSLYRL 600
Db 541 GNTVSDIDPDEEQSTFARSALMRGLGVNSGELYQGLAGVLDAQNTTILSVFVSSLYRL 600
Qy 601 TLLARVHQLTVNELCMYLGSPFNKGTTSLSGSELPRLVILWYQVOTWLTTEAITTEAI 660
Db 601 TLLARVHQLTVNELCMYLGSPFNKGTTSLSGSELPRLVILWYQVOTWLTTEAITTEAI 660
Qy 661 WLICTPFGSGNISPEISNLLNLPSPISSEDMAQSHNRELOAEIILAPTAATLHLASPDMA 720
Db 661 WLICTPFGSGNISPEISNLLNLPSPISSEDMAQSHNRELOAEIILAPTAATLHLASPDMA 720
```

QY 721 RYLLWTDNLPRGGLDIAGFMTLVKESLNANETTQVQFCHVMAQLSLSVQTLLSEAE 780  
 Db 721 RYLLWTDNLPRGGLDIAGFMTLVKESLNANETTQVQFCHVMAQLSLSVQTLLSEAE 780  
 QY 781 LSVLVISGFAVLGAKNQAGOHNDITLFSYRFHOWINGLNGPSDDTLDMRLQOQTLTADR 840  
 Db 781 LSVLVISGFAVLGAKNQAGOHNDITLFSYRFHOWINGLNGPSDDTLDMRLQOQTLTADR 840  
 QY 841 LASVGLDISVMTQAMVAGVNLQOCWQDINTVLQWIDVASALHTMPSVIRTLVNRVYT 900  
 Db 841 LASVGLDISVMTQAMVAGVNLQOCWQDINTVLQWIDVASALHTMPSVIRTLVNRVYT 900  
 QY 901 ALNKAESNLPSWDEWQTLAENNEAGLSLQOQTLADYTAERLSSVLCNFWLANIOPEGVS 960  
 Db 901 ALNKAESNLPSWDEWQTLAENNEAGLSLQOQTLADYTAERLSSVLCNFWLANIOPEGVS 960  
 QY 961 LHSRDDLYSYFLIDNQVSSAIKTTTLAELAIAGIQLYINRALNRPEPNARADVSTROQFTD 1020  
 Db 961 LHSRDDLYSYFLIDNQVSSAIKTTTLAELAIAGIQLYINRALNRPEPNARADVSTROQFTD 1020  
 QY 1021 WTVNRRYSGVSRVLYPENYIDPTQRIQOTRMDELLENISQSLSRDTVEDAPKTY 1080  
 Db 1021 WTVNRRYSGVSRVLYPENYIDPTQRIQOTRMDELLENISQSLSRDTVEDAPKTY 1080  
 QY 1081 LTRPETVADLVKVSAYHDNVNSNTGLTWFGQTRNLPEYWRNVNDISRMQAGELAAANAW 1140  
 Db 1081 LTRPETVADLVKVSAYHDNVNSNTGLTWFGQTRNLPEYWRNVNDISRMQAGELAAANAW 1140  
 QY 1141 KEWKIDTAVNPYKDAIRPVIFRERLHLIWEKEVAKNGDTPVETDRFTLKLAFLRHD 1200  
 Db 1141 KEWKIDTAVNPYKDAIRPVIFRERLHLIWEKEVAKNGDTPVETDRFTLKLAFLRHD 1200  
 QY 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFKYTKGYSYDFGGSNK 1260  
 Db 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFKYTKGYSYDFGGSNK 1260  
 QY 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIHTQGNDLVRKASVFAQDFEVPASLN 1320  
 Db 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIHTQGNDLVRKASVFAQDFEVPASLN 1320  
 QY 1321 MSAIGDLSLTMENGNIPQITSYSSDNLAITLHNAATVRYDGSNVIRNKQISAMKL 1380  
 Db 1321 MSAIGDLSLTMENGNIPQITSYSSDNLAITLHNAATVRYDGSNVIRNKQISAMKL 1380  
 QY 1381 TGVDKSOYGNAFIANTVKHYGYSYDLGGPITVYNTKNTKNIASVQGHLMADYTRRLIL 1440  
 Db 1381 TGVDKSOYGNAFIANTVKHYGYSYDLGGPITVYNTKNTKNIASVQGHLMADYTRRLIL 1440  
 QY 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500  
 Db 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500  
 QY 1501 GWLIDIDGINNTDITKITWAGSKTHFTASDHIALSPANSFDMPTFPKLEIDASSLAF 1560  
 Db 1501 GWLIDIDGINNTDITKITWAGSKTHFTASDHIALSPANSFDMPTFPKLEIDASSLAF 1560  
 QY 1561 TNNIAPLDIVPETKAKORVLGKIKQLSVKRVNVPEDILFLRTHSGAOMQGVTRI 1620  
 Db 1561 TNNIAPLDIVPETKAKORVLGKIKQLSVKRVNVPEDILFLRTHSGAOMQGVTRI 1620  
 QY 1621 RLNTLLASQLVSRANTGIDITLTMETQRLBPPLGEGFFANFVLPKYDPAEHBGDERWFKI 1680  
 Db 1621 RLNTLLASQLVSRANTGIDITLTMETQRLBPPLGEGFFANFVLPKYDPAEHBGDERWFKI 1680  
 QY 1681 HIGNVGNTGRQPYSGMLSDTSETMTLFPVYAEYGMHGVRLGVGYKITDYNTWES 1740  
 Db 1681 HIGNVGNTGRQPYSGMLSDTSETMTLFPVYAEYGMHGVRLGVGYKITDYNTWES 1740  
 QY 1741 AFFYFDETKQOQFVLINDADHDSGMTQOQIVKNIKKYKGLNVSATGYSAPMDFNASAL 1800  
 Db 1741 AFFYFDETKQOQFVLINDADHDSGMTQOQIVKNIKKYKGLNVSATGYSAPMDFNASAL 1800

QY 1801 YWELFYTYPMCFORLLQEKQFDEATOWINVYNPAGYIVNGEITAPWIMNCRPLEETTS 1860  
 Db 1801 YWELFYTYPMCFORLLQEKQFDEATOWINVYNPAGYIVNGEITAPWIMNCRPLEETTS 1860  
 QY 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920  
 Db 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKWYVRT 1920  
 QY 1921 LELLAGDEPEDYGSQQAAPSLSGAASQTVQAAVQODLTMLGRGGVSKNLRANSISLVGLFL 1980  
 Db 1921 LELLAGDEPEDYGSQQAAPSLSGAASQTVQAAVQODLTMLGRGGVSKNLRANSISLVGLFL 1980  
 QY 1981 PEYNPALTDYVQTLRLRLFNLRHNLSDGQPLSLAIYAEPDTPKALLTSMVQASOGGSV 2040  
 Db 1981 PEYNPALTDYVQTLRLRLFNLRHNLSDGQPLSLAIYAEPDTPKALLTSMVQASOGGSV 2040  
 QY 2041 LPGTLISLYRFPVLMERTNLVAQLTOFGTSLLSWAHDDADDELTTLLQQQGMELATQISIR 2100  
 Db 2041 LPGTLISLYRFPVLMERTNLVAQLTOFGTSLLSWAHDDADDELTTLLQQQGMELATQISIR 2100  
 QY 2101 IQORTVDEVDADIIVLAESRERSAQNRLKYQQLYDEEDINHGEQRAMSLDDAAAGQSLAGQ 2160  
 Db 2101 IQORTVDEVDADIIVLAESRERSAQNRLKYQQLYDEEDINHGEQRAMSLDDAAAGQSLAGQ 2160  
 QY 2161 VLSTAEGVADILVNVVFGGLACGSRWGAALRASAVMSLSATASQYSADKISRSEAYRRRR 2220  
 Db 2161 VLSTAEGVADILVNVVFGGLACGSRWGAALRASAVMSLSATASQYSADKISRSEAYRRRR 2220  
 QY 2221 QEWEIFQRDNADGEVQKQMDAQLSLKIRREAAQMOVEYQETOQANTOAOLELLQKFTNKA 2280  
 Db 2221 QEWEIFQRDNADGEVQKQMDAQLSLKIRREAAQMOVEYQETOQANTOAOLELLQKFTNKA 2280  
 QY 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARLRELTONGVTFFIRGGAWNGTTAGLMAGET 2340  
 Db 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARLRELTONGVTFFIRGGAWNGTTAGLMAGET 2340  
 QY 2341 LLLNLAEMEKVWLBRRDERALEVTRTVSLAQYQALSSDNFNLTEKLTQFLREGKGNVGAS 2400  
 Db 2341 LLLNLAEMEKVWLBRRDERALEVTRTVSLAQYQALSSDNFNLTEKLTQFLREGKGNVGAS 2400  
 QY 2401 GNEUKLSNRQIEASVRLSDLKI FSDYPESLGNTRQLQKQSVTLTPALVGPYEDIRAVLNG 2460  
 Db 2401 GNEUKLSNRQIEASVRLSDLKI FSDYPESLGNTRQLQKQSVTLTPALVGPYEDIRAVLNG 2460  
 QY 2461 GSIWMPRGCSAIALSHGVNDSGQFMDNFNDNRYLPFEGISVNDSGSLTSLFPDATDROKA 2520  
 Db 2461 GSIWMPRGCSAIALSHGVNDSGQFMDNFNDNRYLPFEGISVNDSGSLTSLFPDATDROKA 2520  
 QY 2521 LLESLSDIILHIRYTIRS 2538  
 Db 2521 LLESLSDIILHIRYTIRS 2538

RESULT 4  
 US-10-754-115-34  
 ; Sequence 34, Application US/10754115  
 ; Publication No. US20040208907A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Schleper, Amanda  
 ; APPLICANT: Bevan, Scott  
 ; APPLICANT: Bintirim, Scott  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Ni, Weiting  
 ; APPLICANT: Zhu, Baolong  
 ; APPLICANT: Merio, Don  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Meade, Thomas  
 ; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
 ; FILE REFERENCE: DAS-104XCI  
 ; CURRENT APPLICATION NUMBER: US/10/754,115  
 ; CURRENT FILING DATE: 2004-01-07

```
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 2538
; TYPE: PRF
; ORGANISM: Xenorhabdus nematophilus
; us-10-754-115-34

Query Match      100.0%; Score 13043; DB 4; Length 2538;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYSTAVLUNKISPRDGGTMTLADQLVLSFELARKIFDDQLSWGAEARHLVHETIEQKNN 60
Db 1 MYSTAVLUNKISPRDGGTMTLADQLVLSFELARKIFDDQLSWGAEARHLVHETIEQKNN 60

Qy 61 RLLEARIPTFRANPOLSGAIRLGIERDSVRSYDEMFGARSSSVKPGSVASMPSPAGYLT 120
Db 61 RLLEARIPTFRANPOLSGAIRLGIERDSVRSYDEMFGARSSSVKPGSVASMPSPAGYLT 120

Qy 121 ELYREAKDLHPSSAYHLDNRRLADLTLSQSNMDEISLTLSNELLLEHITRKTTGGD 180
Db 121 ELYREAKDLHPSSAYHLDNRRLADLTLSQSNMDEISLTLSNELLLEHITRKTTGGD 180

Qy 181 SDALMESLSTVROAIDTPYHOPYETIRQVIMTHDSTLSALSRNPEVMQAEASLLAILA 240
Db 181 SDALMESLSTVROAIDTPYHOPYETIRQVIMTHDSTLSALSRNPEVMQAEASLLAILA 240

Qy 241 NISPELVNLTETETEKADALFAQNFSENITPENFASQSWIAKYIGLSEVOKYLGML 300
Db 241 NISPELVNLTETETEKADALFAQNFSENITPENFASQSWIAKYIGLSEVOKYLGML 300

Qy 301 QNGYSDSTSAYVDNISITGLVNNESKLEAYKIRTKVTDYDKNINYPDLMEYEGNNQFFIR 360
Db 301 QNGYSDSTSAYVDNISITGLVNNESKLEAYKIRTKVTDYDKNINYPDLMEYEGNNQFFIR 360

Qy 361 ANPKVSRFEGATLRKNAQPSGIVGSLGPLIANTNFNFKSNYLSNISDSBYKNGVKIYAYRY 420
Db 361 ANPKVSRFEGATLRKNAQPSGIVGSLGPLIANTNFNFKSNYLSNISDSBYKNGVKIYAYRY 420

Qy 421 TSSTSATNQGGGIFTFESYPLTIFALKLNKAIKRLTSGLSNLEQTTIVRSDNAQGIIND 480
Db 421 TSSTSATNQGGGIFTFESYPLTIFALKLNKAIKRLTSGLSNLEQTTIVRSDNAQGIIND 480

Qy 481 SVLTKVFTFLFYSHRYALSFDDAQVLANGSVINQYADDDSVSHFNRLFNTPLKGIPEAD 540
Db 481 SVLTKVFTFLFYSHRYALSFDDAQVLANGSVINQYADDDSVSHFNRLFNTPLKGIPEAD 540

Qy 541 GNTVSIIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600
Db 541 GNTVSIIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600

Qy 601 TLLARVHQLTVNELCMLVGLSPFKGKTTFASLSSELPLRLVTLVQVTCMLTEASITTEAI 660
Db 601 TLLARVHQLTVNELCMLVGLSPFKGKTTFASLSSELPLRLVTLVQVTCMLTEASITTEAI 660

Qy 661 WLLCTPESGNIISPEISNLLNLRPSISEDMAQSHNRELOAEIILAPFTAAHLHASPMDMA 720
Db 661 WLLCTPESGNIISPEISNLLNLRPSISEDMAQSHNRELOAEIILAPFTAAHLHASPMDMA 720

Qy 721 RYILLWTNLRPPGLDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780
Db 721 RYILLWTNLRPPGLDIAGFMTLVLKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780

Qy 781 LSVLVISGFAVLGAQNQAGQHNIDTLFSLYRFHQWINGLNGPSSDITLMDLRQOQLTADR 840
Db 781 LSVLVISGFAVLGAQNQAGQHNIDTLFSLYRFHQWINGLNGPSSDITLMDLRQOQLTADR 840

Qy 841 LASVMGLDISMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900
Db 841 LASVMGLDISMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900

Qy 901 ALKAESNLPSWDEWQTLAENMEAGLSTQAOATLADYTAERLSSVLNCWFLANIQPEGVS 960
Db 901 ALKAESNLPSWDEWQTLAENMEAGLSTQAOATLADYTAERLSSVLNCWFLANIQPEGVS 960

Qy 961 LHSRDDLYSYFLIDNQVSSAIKTRRLAEAIAGIQLYINRALNRIEPPNARADVSTROFFTD 1020
Db 961 LHSRDDLYSYFLIDNQVSSAIKTRRLAEAIAGIQLYINRALNRIEPPNARADVSTROFFTD 1020

Qy 1021 WTNNRYSTWGVSRSLVYPENYIDPTQRIQOTRMDELLENISQSKLSDRTVEDAFKTY 1080
Db 1021 WTNNRYSTWGVSRSLVYPENYIDPTQRIQOTRMDELLENISQSKLSDRTVEDAFKTY 1080

Qy 1081 LTRPETVADLKVSAYHDNNSNTGLTFVQGTRENLPEYYWRNVDISRWQAGELAANAW 1140
Db 1081 LTRPETVADLKVSAYHDNNSNTGLTFVQGTRENLPEYYWRNVDISRWQAGELAANAW 1140

Qy 1141 KEWTKIDTAVNPYKDAIRPVI FRERLHLI WVEKEEVAKNGTDPVETYDRFTLKLAFRLHD 1200
Db 1141 KEWTKIDTAVNPYKDAIRPVI FRERLHLI WVEKEEVAKNGTDPVETYDRFTLKLAFRLHD 1200

Qy 1201 GWSAPWSYDITTOVEAVTDKPDTERLALAAAGFQGEDTLLVPVYKTGKSYPFGGSNK 1260
Db 1201 GWSAPWSYDITTOVEAVTDKPDTERLALAAAGFQGEDTLLVPVYKTGKSYPFGGSNK 1260

Qy 1261 NVAGMTIYGDGSPKKMENTALSRYSQKNTPDIIHTQGNLVRKASVRFAPQDPVPASLN 1320
Db 1261 NVAGMTIYGDGSPKKMENTALSRYSQKNTPDIIHTQGNLVRKASVRFAPQDPVPASLN 1320

Qy 1321 MGSAIGDSDSLTVMENGNIPOITTSKYSSDNLAITLHNAAFVRYDGSNGVIRNKQISAMKL 1380
Db 1321 MGSAIGDSDSLTVMENGNIPOITTSKYSSDNLAITLHNAAFVRYDGSNGVIRNKQISAMKL 1380

Qy 1381 TGVDGKSYQGNAFIANTVVKHYGYSDLGGPITVYNTKQNIASVQGHLMNADYTRLLIL 1440
Db 1381 TGVDGKSYQGNAFIANTVVKHYGYSDLGGPITVYNTKQNIASVQGHLMNADYTRLLIL 1440

Qy 1441 TPVENNYARLPFPFPSPNTILNTVTVGSKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500
Db 1441 TPVENNYARLPFPFPSPNTILNTVTVGSKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500

Qy 1501 GWLDDITGINNTDIKITVMAGSKTHFTASDHIAASLPANGSPDAMPYTFKPLEIDASSLAF 1560
Db 1501 GWLDDITGINNTDIKITVMAGSKTHFTASDHIAASLPANGSPDAMPYTFKPLEIDASSLAF 1560

Qy 1561 TNNIAPLDIIVPETKAKDGRVLGKIQTLSVKRVNYPEDILFLRETHSGAQYMLGVYRI 1620
Db 1561 TNNIAPLDIIVPETKAKDGRVLGKIQTLSVKRVNYPEDILFLRETHSGAQYMLGVYRI 1620

Qy 1621 RLNTLLASQLVSRANTGIDITLTMETORLPPEPLGEGFFANFVLPKYDPAEHGDERWPKI 1680
Db 1621 RLNTLLASQLVSRANTGIDITLTMETORLPPEPLGEGFFANFVLPKYDPAEHGDERWPKI 1680

Qy 1681 HIGNVGNTGRQPYYSGLMDSSTSMTLFVPAEYGMHEGVRGLGVGYQKITVDNTWES 1740
Db 1681 HIGNVGNTGRQPYYSGLMDSSTSMTLFVPAEYGMHEGVRGLGVGYQKITVDNTWES 1740

Qy 1741 AFFYFDETKOQFVLINDADHDSGMTQOGIVKNIKKYKGLNVSTATGYSAPMDNFSASAL 1800
Db 1741 AFFYFDETKOQFVLINDADHDSGMTQOGIVKNIKKYKGLNVSTATGYSAPMDNFSASAL 1800

Qy 1801 YYWELFYVTPMWCFORLLQEKQFDEATQWVNVNYPAGIYVNGEIAPIWNCRPLEETTS 1860
Db 1801 YYWELFYVTPMWCFORLLQEKQFDEATQWVNVNYPAGIYVNGEIAPIWNCRPLEETTS 1860

Qy 1861 WNAFLDAIDPDVAQNDPMHYKIATFMRLDQILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNAFLDAIDPDVAQNDPMHYKIATFMRLDQILRGDMAYRELTRDALNEAKWYVRT 1920

Qy 1921 LELLGDPEPDYGSQWAAAPSLSGAASQTVQAAYOODLTMLGRGVSKNLRANSIUVGLFL 1980
Db 1921 LELLGDPEPDYGSQWAAAPSLSGAASQTVQAAYOODLTMLGRGVSKNLRANSIUVGLFL 1980
```

```
QY 1981 PEYNALTDYWTLLRLFLNLRHNSIDGQPLSLAIYAEPDTPKALLTSMVQASQGSVAV 2040
DB 1981 PEYNALTDYWTLLRLFLNLRHNSIDGQPLSLAIYAEPDTPKALLTSMVQASQGSVAV 2040
QY 2041 LPGTLISLYRFPVMLRTRNLVAQLTOFGTSLISMAEHDDADELTLLLOOQGMELATQSIR 2100
DB 2041 LPGTLISLYRFPVMLRTRNLVAQLTOFGTSLISMAEHDDADELTLLLOOQGMELATQSIR 2100
QY 2101 IQORTVDEVDADIAVLAESRSQAQRLEKYQOLYDEDINHGEQRAMSLDDAAAGQSLAQ 2160
DB 2101 IQORTVDEVDADIAVLAESRSQAQRLEKYQOLYDEDINHGEQRAMSLDDAAAGQSLAQ 2160
QY 2161 VLSIAGVADLVPNVFLGACGSRWGAAALRASASVMSLSATASQYSADKISRSEAYRRRR 2220
DB 2161 VLSIAGVADLVPNVFLGACGSRWGAAALRASASVMSLSATASQYSADKISRSEAYRRRR 2220
QY 2221 QEWELQRNADGEVKOMDAQLESKIRREAAQMOVETOQAHTQAOLELLQRKFTNKA 2280
DB 2221 QEWELQRNADGEVKOMDAQLESKIRREAAQMOVETOQAHTQAOLELLQRKFTNKA 2280
QY 2281 LYSWMRGKLSAIYYOFFDLTOSFCLMAQEARRELTONGVTFIRGAWNGTTAGLMAGET 2340
DB 2281 LYSWMRGKLSAIYYOFFDLTOSFCLMAQEARRELTONGVTFIRGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNTEKLTQFLREGKGNVGAS 2400
DB 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNTEKLTQFLREGKGNVGAS 2400
QY 2401 GNEKLSNRQTEASVRLSDIKIFSDYPESIGNTRQOLKQVSTLPAIVGYPEDIRAVLNYG 2460
DB 2401 GNEKLSNRQTEASVRLSDIKIFSDYPESIGNTRQOLKQVSTLPAIVGYPEDIRAVLNYG 2460
QY 2461 GSIVMPCGSAIALSHGVNDGSGQFMDLNDNRYLPFEGISVNDGSGSLTSPDPAIDROKA 2520
DB 2461 GSIVMPCGSAIALSHGVNDGSGQFMDLNDNRYLPFEGISVNDGSGSLTSPDPAIDROKA 2520
QY 2521 LLESLSDIILHRYTIRS 2538
DB 2521 LLESLSDIILHRYTIRS 2538
```

## RESULT 5

```
US-10-753-901-14
; Sequence 14, Application US/10753901
; Publication No. US20040194164A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Larrinua, Ignacio
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Schafer, Barry
; APPLICANT: Bevan, Scott
; APPLICANT: Young, Scott
; APPLICANT: Guo, Lining
; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control
; FILE REFERENCE: DAS-105X
; CURRENT APPLICATION NUMBER: US/10/753,901
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,717
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-753-901-14
```

```
Query Match 42.1%; Score 5496.5; DB 4; Length 2523;
Best Local Similarity 44.7%; Pred. No. 0;
Matches 1168; Conservative 436; Mismatches 833; Indels 175; Gaps 36;
```

QY 7 LLNKISPTRDGGTWTADLQVLSFSELKRI FDDQLSWGEARHLYHETIEQKKNRLLEAR 66

```
DB 7 LLDKINRKRSGDTLLLTNISFMSFEFRHRTSGTLTWRETDFLYQQAQBSKQNKLEBLR 66
QY 67 IFTANPOLSCAIRGLGIERDSVSRSDMGARSSSVKPGSVASMPFAGVLTLEYREA 126
DB 67 ILSRANPOLANTWNITPSTLNNSYNWFGRAHRFVKPGSIASIFSPAAYLTLEYREA 126
QY 127 KDLHSSAYHLNRRPDLADLTLSQSNMDTEISTLTLSNELLLEHITRKTGGSDALME 186
DB 127 KDFPDNSQYHLNKRPPDIASLALTQNMDEISTLSNELLHNIQTLEKTDYGVYMK 186
QY 187 SLSTYRQADTPYQHPYETIROVIMTHDSTLSALSRNPVWQAGASILLAILANISEL 246
DB 187 MLSTYRQGMTPYHLPYESARQALLQDKNITAFSRNTDVAELMDPTSLLAIKTDISSEL 246
QY 247 YNILTETEKNADALPAQNF--SENITPENFASOSWIAKYKYGLELSEVQVYKQLMQLNGYS 305
DB 247 YQILVEEITPENSTELMKCNFTDDVL--IFKSVASLARYYDLSVDELSLFW--LSFGCK 303
QY 306 DSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINYPFDLMYEGNQQFFIRANPKV 365
DB 304 NTNQYKNEQLITLVNDGNDTATARLIKTRKDPYDSHLNAELIPIKENEY--KYNFSV 361
QY 366 SREPGATL--RKNAGPSGIVGSLSGPLIANTNPKSNVLSISDSEYKNGVKIAYRYTS 422
DB 362 KKTBPDLDFRLQNGKEYIYQDNFVPIANTHY--SIPIKLTTEQITNGITLRLWRVKP 419
QY 423 STSATNCGGFTFESVPLTIPALKNAIKRLCTLSGLSPNELOTVIRSDNAQGIINDSV 482
DB 420 NPSDAINNAYFKWMEFGDIFLLKNAIKRLYKATGISPEDIQVIESIYDDLTIDSNV 479
QY 483 LTKVFTYLFYSHRYALSPDDAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGIPEADGN 542
DB 480 LGKLPYQYVQYHYNISVS DALVLSHSDISQYSTKQPSHFTILFNTPELLNGQESADNT 539
QY 543 TVSIDPBEQSTFARSALMRGLGVNSGELYQGLKAGLVLDQNTITLSVFISSLYRLTL 602
DB 540 KLDLTPGESKNHFYLGIMKRAFRVNDTELYTLWKLANG--GTNPEFMCSEIENLSLYRVL 598
QY 603 LARVHOLTVELCMLYGLSPENGGTKTASLSGGLPRLVIMLYQVTVQWLTAEITTEAIWL 662
DB 599 LADIIHLTVNELSMLLSVSPVNTKIALFSDTALTQLISFLPQCTQWLTQKWSVDVFL 658
QY 663 LCTPEFGNISPEISNLNNLRPSISEDMAQSHNRELOAEILAPFIAATLHLPASDMARY 722
DB 659 MTDNYSVLTPTDIENLITLSNGLS--TSLGDELIRAA--AAPLIAASIQMSAKTAET 715
QY 723 ILLWTDNLRPGGLDIAGFMTLVLESNANETTVQVCHYMAQLSLSVQTLRLSEAEIS 782
DB 716 ILLWINGIKPQGLTFDDFMIIAANRDRSENETSNMVAFCQVLGQLSLIVRNIGUSELT 775
QY 783 VLVISGPAVLGAKNOP-----AGQHNIOTLFSLYRFHQWINGLGNPGSDTLDMLRQQ 834
DB 776 LLV-----TKPEKFQSEITTAHQDHLPTLQALTFHAVIMRCGSYATEILLTALBLG 825
QY 835 TLTDADRLASVGLDISMVTQAM--VSAGVNOQCQDINTVQLQWIDVASALHTWPSVIRT 892
DB 826 ALTRAQLAVALKFDQVVTQALQQTGLGVNTFTNWRITIDVTQLQWIDVAATLGTIPDGVA 885
QY 893 LVNTRYVTALKAESNLPMSWDEWQTLAENMEAGLSQQAOTLADYTAERLSSVLCNWF 952
DB 886 LILKLYI---GEPETPMPTFDDWQAATLLQAGLNSQSQSDQLQAWLDEATTATAASYIK 942
QY 953 NIOPEGVLSHRDLYSYFLIDNQVSSAIKTRTLAEAIAGQLYINRNLNRIEPPNARADV 1012
DB 943 NSAPQ--QIKSRDELYSYLLIDNQVSAQVKTTRVAEAIASQLYVNRALNVEGKVP 1000
QY 1013 STROFFTDW--TVNNRYSTWGCVSRLVYPENYIDPTQRIGQTRMDELLENISQSKLSRD 1071
DB 1001 KTRQFFCDWETYNRRYSTWAGVSELYVYPENYIDPTIRIGQTRGMNNLLQOLSQSLNID 1060
QY 1072 TVEDAPKTYLTRFETVADLKVVSAYHDVNSNTGLTWFGVGTRENLPEYKRVNDISRMQ 1131
```



Db 1061 TVEDSPKNYLTAFSDVANLQVISHVDISINVEGITYLIGYSQTEPRYYWRNVHDHOKCQ 1120  
Qy 1132 AGEIAAANAKWTKIDTAVNPYKDAIRPVI PRERHLIWEKEVAKNGTDPVETYDRFT 1191  
Db 1121 HGQFAANAWGKKIEIPINWQENIRPVIYKSRLLYLWLQKELKNESEDKIDITDIYI 1180  
Qy 1192 LKLAFLRHGDSWSPWSYDITTOVEAVTDKPKDTERLALASGFGQEDTLVAVVYKTKGS 1251  
Db 1181 LKLSHIRYDGSWSPFNFNVDKLENLKKAKS---IGMYCSDYKEDVILVYFHEKKDN 1237  
Qy 1252 YSDFGSKNKNVAGMTIYGDGSKKMENTALSRYSOLKNTFDIITQGNLDLVRKASYRPAQ 1311  
Db 1238 YS--FNSLPAREGMTINPDWTLISILTENDLD--AIVKSTLSELDTRTE--YKVNQFAT 1290  
Qy 1312 DF-----EVPASLNMGSAGIG---DDSLTVVNGNI-----PQTSKYSSN----- 1349  
Db 1291 DYLAEBKESITTKNKLASFTGNI FDLSYISPNGHINLTFNPSMEINFPSKGNINYDEVKY 1350  
Qy 1350 -LAITLHNAATVVRDGSNGVIRNKQISAMKLTGVDSKSOYGNAFIANTVKKHYGGYSDL 1408  
Db 1351 LLSMVEDETVILFDYDRHEML-----GKEE-----EYFHYG----- 1382  
Qy 1409 GGPITVYNTKNTKNTYASVOGHLMNADYTRRL-----ILTPVENNYIARLFEEF 1454  
Db 1393 -----TLDFIISID--LKNAEYFVLMHLRTKEKI PRKSEIGVGINVDYESNDAEF 1431  
Qy 1455 PFSPTNLTN-----TVPTVGSNTKSDPK-----KCSYAVDGNNS 1488  
Db 1432 KLDNTNVLWDKNTGVWHTICESEFTNDVSIINNMGNI AALFLREDPCVYLCSATD---- 1487  
Qy 1489 QGFOIFSSYQSSGMLDIDTGINNTDITKITVMAGSKTHT-FTASDHASLSPANSFDMAPYT 1547  
Db 1488 --IKIASSNIEQ-----IQDKNISFLKNGSDILVELNABEDHVASKPSHESDPWYD 1537  
Qy 1548 FKPLEIDASSLAFTNIIAPLDIVFETKAKGRVLG-KIKOTLSVKRVNYPNEDILFLRET 1606  
Db 1538 FNQVKVDIEGYDI-----PLVSEFIKQPDGGYNDIVIESPIHIKLSKDSNTSLHKM 1592  
Qy 1607 HSGAQYQMGVYRIRLNTLLASOLVS RANTGIDITLMTETORLBPPLGEGFFANFVLPK 1666  
Db 1593 PSQTQYMQIGYRTRLNTLFRKLABRANIGIDNVLSTMETQNLPEPQGEGFYATFKLPP 1652  
Qy 1667 YDPAEHGDERFKHIGVNGTGRQPYSGMLSDTSETSLFVPYAEYGMHEGVRIG 1726  
Db 1653 YNKEHGDERFKHIGNIDNSARQPYEGMLSDI-ETTTLFVPYAKGYIIEGVRIG 1711  
Qy 1727 VGYQKITDNTWESAFFYDETKQOQVFLINDADHDSGMTQOQIVKNIKKYKGLNVSAT 1786  
Db 1712 VGYKKIYDKSWESAFFYDETKNQIFINDADHDSGMTQOQIVKNIKKYKGFHVVVMK 1771  
Qy 1787 GYSAPMDPNSALYYWELFYTPNMCORLLOEQFDEATQWYVYVNPAGYVNGEIA 1846  
Db 1772 NNTPEMPDNGANAIYFWELFYTPNMFVQRLLOEQNFTESTRWLYIWNPAQSYQGSQM 1831  
Qy 1847 PWIKNCRPLEETSNANPLDAIDPDVAONDPMHYKIATFMRLLDOLILRGDMAYRELT 1906  
Db 1832 DYNNVNRPLEEDTSNANPLSDVDPDVAQHDPMHYKVATFMKLDLLITRGDSAYRQLE 1891  
Qy 1907 RDALNEAKWYVYRTLELLGDEPDYDGSQOQAPSLSGAASQTVQAAYQODLTMLGRGVGS 1966  
Db 1892 RDTLNEAKWYVYQALTLTGDEPFYFLSDNDWSEFPLEEASQTMRRHHYQHKMLQRQRAAL 1951  
Qy 1967 KNLRTANSVLGLFLPEYNPALTDYQOTLRLFLNRLNLSIDGQPLSLAIYAEPTDPKAL 2026  
Db 1952 PTKRTANSLTALFLPQINKLQGYWQTLTQRLYNLRLNLTIDGQPLSLSLYATPADPSML 2011  
Qy 2027 LTSVMQASOGGSVLPGLTSLIYRFPVLMLETRNLVAQLTQFCTSLLSMAEHDDADDELFTL 2086  
Db 2012 LSAAITASOGGDDLPHAWPMYRFPVLIENAKWGSQILIQFNTLLSITERQDAEALAEI 2071  
Qy 2087 LLOQGMELATOSIRIQORTVDEVDADI AVIABRSRQAURLKYEKOOLYDEIDINHGQRAM 2146  
Db 2072 LQOQSELALQSIKMQDMAEIDADKLALQESRHGAQSRFDSFNTLYDEVDNAGEKQAM 2131

Qy 2147 SLILDAAGQSLAGQVLTSIAGSVADLVNVPFGLACGSRGWAALRASASVMSLSATASOYS 2206  
Db 2132 DLYLSSSVLSTSGTALHMAAADLVNIIYGFVGGSRFGALFNASAIIGTISASATRIA 2191  
Qy 2207 ADKISRSEAYRRRRQEWIEIQORDNADGEVKQVDAQLESKIRREAAQMOVYQETQQAHTQ 2266  
Db 2192 ADKISQSEIYRRRRQEWIEIQORNAEAEIKOIDAQLATLAVRREAAVLQKNYLETQQAQQT 2251  
Qy 2267 AQLLELQKNTKALYSWMRGKLSAIYYQFEDLTQSCFMAQEAALRELTDTNGVTFIRGG 2326  
Db 2252 AQLAFLQSKFTSNAALYNLGRSLSAIYYQFYLAVSLCLMABQTYQYELANNAAHFKPG 2311  
Qy 2327 AWNGTTAGLMAGETLLINLAEMEKVLMLEDERALEVTRTVSLAQFYQALSSDNFNLTTELK 2386  
Db 2312 AMHGTYAGLAGETMLNLAQMEKSYLEKDERALEVTRTVSLAEVYAGLTENSILKDKV 2371  
Qy 2387 TQFLREKGNVANGSGBNELKLSNRQIEASVRLSDUKIFSDYSPESLGNTRQLKQVSVTLPAL 2446  
Db 2372 TELVNAGEGSGAGTTLNGINVEGTQLQASLSKLSDLNIATDYPDGLGNTRRIKOISVTLPAL 2431  
Qy 2447 VGPYEDTRAVLNYGGSIWMPRGCSAIALSHGVNDSGQFMPLDFNDSRYLPPFGISVNDSGS 2506  
Db 2432 LGPYQDVRALLSYGGSTMMPRGCKAIAISHGMNDSGQFQMDFNDAKYLPPFEGLPVADTGT 2491  
Qy 2507 LTLSPDPATDROKALLBSLSDIILHRYTIRS 2538  
Db 2492 LTLSPGIGSKQKSLLSLSDIILHRYTIRS 2523  
RESULT 6  
US-10-754-115-14  
; Sequence 14, Application US/10754115  
; Publication No. US20040208907A1  
; GENERAL INFORMATION:  
; APPLICANT: Hey, Timothy  
; APPLICANT: Schleper, Amanda  
; APPLICANT: Bevan, Scott  
; APPLICANT: Birtim, Scott  
; APPLICANT: Mitchell, Jon  
; APPLICANT: Li, Ze Sheng  
; APPLICANT: Ni, Weiting  
; APPLICANT: Zhu, Baolong  
; APPLICANT: Merlo, Don  
; APPLICANT: Apel-Birkhold, Patricia  
; APPLICANT: Meade, Thomas  
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control  
; FILE REFERENCE: DAS-104XC1  
; CURRENT APPLICATION NUMBER: US/10/754,115  
; CURRENT FILING DATE: 2004-01-07  
; PRIOR APPLICATION NUMBER: US 60/441,723  
; PRIOR FILING DATE: 2003-01-21  
; NUMBER OF SEQ ID NOS: 64  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 14  
; LENGTH: 2523  
; TYPE: PRT  
; ORGANISM: Xenorhabdus nematophilus  
US-10-754-115-14  
Query Match 42.1%; Score 5496.5; DB 4; Length 2523;  
Best Local Similarity 44.7%; Pred. No. 0;  
Matches 1168; Conservative 436; Mismatches 833; Indels 175; Gaps 36;  
Qy 7 LNKISPTDQQTWTLADQLVLSSELSRKIFDDQLSWGEARHLHYETIEOKKNLEEAR 66  
Db 7 LLDKINKRSQDITLLTNISFMSFSEPHRTSGTLTWRETDFLYQQAQHSQKQKLELR 66  
Qy 67 IFRANPOLSCAIRLIGTERDSVRSYDEMFCAGSSSFVKPGSVASMFSPAGYLTLYREA 126  
Db 67 ILSRANPOLANTTNLTPTSLNNSYNSWFGRAHFRVKPGSIASIFSFAAYLTLYREA 126  
Qy 127 KDLHFSSAYHLNRRPDLADLTLSQSNMOTETISTLTLSNLELLEHITRKTGGSDALME 186









		SEQUENCE CHARACTERISTICS:			
		LENGTH: 2516 amino acids			
		TYPE: amino acids			
		TOPOLOGY: linear			
		MOLECULE TYPE: protein			
		US-10-262-794A-47			
		Query Match 38.7%; Score 5043; DB 4; Length 2516;			
		Best Local Similarity 42.6%; Pred. No. 0;			
		Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;			
Qy	22	LADQLYLSFSLRIRKIFDDQLSWGBARHLYHETIBQKNNRLLLEARIETFRANPOLSGAIRL	81	885	QQLNVAPOGVSAVLGLDYIQSMKET-----PTYAQWENAGVLTAGLNSQQANTLHAFLE
Db	21	LTDLSHSFNFPROQVSEHLSWSETHDYDHAQQAQKNDRLYEAILKRAPOLQNAVHL	80	941	RLSSVLCNWFLANIQPEGVLSHSRDDLYSYFLDNDVOUSSAKTRTLAEATAGIOLYNRA
Qy	82	GIERDSVSR-SYDEMFGARSSFFVKPSGVSAMFSPAGYLTLYREAKDLHPSSAYHLDN	140	941	SRSAAALSTYYIRQVAKAAAAIKSRDDLYQYLLDNDQVSAAKTRTRIAEASIASIOLYNRA
Db	81	AIALPNAELIGYNNQFSGRASQYVAGTVSMFSPAAYLTLYREARNLHASDSVYLDL	140	1001	LNRIEPNARADVSTRQFFTDW-TVNNRYSTGGVSRVLYYPENYIDPTQIGOTRMDEL
Qy	141	RRPLADLTLSQNMDEITSLTSLNELLLEHI-TRKTGGSDALMESLSTYRQAIDTPY	199	1001	LENVEENANSGVLSRQFFIDWDKYNKEYSTWAGVSQVLYYPENYIDPTQIGOTKMDAL
Db	141	RRPLKSMALSQQNDIELSTLSNELLLESIKTESKLENYTKVMEMLSTFRPSGAPPY	200	1060	LENISQSKLSDRVEDAFKTYLRFETVADLKVSAHDNNSNTGLTWFGQTRENLUPE
Qy	200	HOPIETIROVTMTDSTLSALSRNPEVMGQAGASLLAILANISPELYNLTETIEKNA	259	1061	LQSVSQSLNADTVEDAFMSYLTSEFQVANLKVISAYHDNNDQGLTYFGLSETDAGE
Db	201	HDAYENVREVLOQDPLGLEQLNASPAIAGLMHQSALLGINASISPELFNLTETIEGNA	260	1120	YYRNVNDVSRMQAGELAAWAKWTKIDTAVNPVKDAIRPVIPIERLHLHVEKEEVAK-
Qy	260	DALPAQNFSENITPENFASQSWIAKYGLEGLESEVQKYLGMLOK-GYSDSTSAYVDNISTG	318	1121	YYRNVSDHSPKDFNDGKFAANAWSEWHKIDCPINPYKSTIRPVIYKSRLLYLLWLEQKETKQ
Db	261	EELYKKNFG-NIEPASLAMPYLRKYNNLSDEELSQPIGKASNFQQEYSN---NQLITP	316	1179	--NGTD--PVETDYDRFTLKLAFRLRHGDSWSPASYSYDITTOVEAVTDDKPPDTERLALAASG
Qy	319	LNVNNEKLEAYKTR-VKTDYDKNINYPDLMEYEGNQPPIRANFKVSRFEGAT-----	372	1181	TGNSKDGQYQETDYRYELKLAHRYDGTWPTITFDVNNKKSIEL--KLEKNRAPGLYCAG
Db	317	VNNSSDGTGVKYRITREYTNAYQMDVELFP--FGEN---YRLDYFKPNFYNASYLSIK	371	1235	FOGEDTLVLYVYTKGYSDFGSGNKNVAGWTIYVGDSPFKKMENTALSRSQSLKNTFDII
Qy	373	-----LRKNAGPSGIVSGSLPLIANTNFKSNVLSNISSEYKNGVKIYAVRYTSTGS	425	1239	YQGEDTLVLMFYNQDRLDSY--KNASMQGLYIPADMASKDMTPEQSNVYRD--NSYQQF
Db	372	LNDKRELVRTGAQO-----VNIEYSANITLNTAD--ISQPEIGLTVLPSSG	418	1295	HTQGNDAVRKASVYRPAQDFEVPASLWNGS--AIGDDSLTYMENGNIPOITSKYSSDNLAI
Qy	426	ATNOGGGIFTPESEYPLTIFALKANKAIRLCITUSGLSPNELTIVRSNQAQIINDSVLTK	485	1295	DTNN---VRRVNNRYAEDEIPSSVSRKDYGMWDYLYSMVYNGDITPIYNKAASSDUKI
Db	419	WA-YAAAKFTVEEYNOYSEFLKLNKAIKRLSRATELSPTILEGIVRSVNLQDINTDVLGK	477	1353	-----TLNNAFTVRYDGSNVIRNKQISAMKLTGVDSKQSYGNAPFIANTVKHYGGY
Qy	486	VFTYLFYSHRVALSFDDAQLNGSVINQYADDSVSHENLFTNPPKLGKIFPEADGNTVS	545	1352	YISPKLRIHNG-----YEQG-----KENQCNLMNKG-----KLGDKFIV-----Y
Db	478	VFLTKYMYQRYAHAEATILICNAPISQSRSDYDNPQSDRFLNTPLLNGQFSTGDEID	537	1406	SDLG-GPITVYNTKKNY-IASVQGHLMNADYTRRLIITPVENNYIARLFEF-PPSPNTIL
Qy	546	IDPBEQSTFARSALMRGLGVNSGELYQLGKLAGVDAQNTITLSVFISSLYRLTLIAR	605	1389	TSLGVNPNNSNKLMPYVYQYSGNTSGLNQGR--LLFHRDITFTYPSKEAWIPAKERSLT
Db	538	LN-SGSTGDWRKTLKRAFNDIDVSLFLLAKITDHNKDGKIKNNLKNLSNLYIGKLLAD	596	1463	NTVFTVGS-----NKTSPFKCSYAVDGNNSOGFOIPSSYQSSGMLDIDTGINNTDIK
Qy	606	VHQLTVNELCMLYGLSPNGKTTAS-LSSGELPRVIVLYQVOTWLTEAETITRAIWLCC	664	1447	NONAIGDDVATOSLNKPDLLKQYIFMTD---SG-----TATDVSGPVEINTALSPAKVQ
Db	597	IHQLTIDELDLLL-IAVGEKTNLSAISDKQLATLIRKLNITITSWLHTQKWSVQLPIMT	655	1516	ITVMAGSKTHTFTASDHASLIPANSFDAMPYFKPLEIDASSLAFTNNIAPLDIVFFTKA
Qy	665	TPFSGNISPRISNLLNLRPSI---SEDMAQSHNRELOAAILAPFIATLHLASPDWAR	721	1500	IIVKAGGKEQTFTADKDVSIQSPSPFDEMYQFNALSIDGSLFINNSASIDVTFTAPA
Db	656	STSNKTLTPRIKNLLDTVYHGLQFDKDKAD-----LLHVMAPYIAATLQLSENVAH	709	1576	KDGRVLG-----KIKQTLVSKRVNTNPEDILFLRETHSGAQVMQLGVRIRLNTLLASQLV
Qy	722	YILLWTDNLRPGGLDIAGFMT-----LVLKESLNANETQO-----LVQFCHVMAQLSLS	770	1560	EDGRKLGVESFSIPVTLKV-----STDNALTLLHNENGAQYMQWQSYRTRTLTLFARQLV
Db	710	SVLLWADKLQPD-----GAMTAKEFWDLNLTYPGSSSEAVEYTOEHIVQYQALQALEMV	765	1632	SRANTGIDTILTMETQRLPEPPLGEGFFANFVLPKYPDAEHGDERWFKIHIGNVGGNTGR
Qy	771	VQTLRLSBAELSVLVISGFAVLGAKNQAGQHNIDTLPSLRFHQWINGLGNPQSDTLDM	830	1615	ARATGIDTILTMETQRLPEPPLGEGFFANFVLPKYPDAEHGDERWFKIHIGNVGGNTGR
Db	766	YHSTGINENAPRLFTVKE-EMFGAATGAAPAHADALSILMLTRFADWNVALGEKASSVLAA	824	1692	QPYTSGMLSDTSETSMTLF-----VPAEYTHMEHGVRLGVGYQKITYDNTWESAFFYFD
Qy	831	LROQTLTADRLASVMGLDISWYQAMVSA-----GVNQLQODINTVLQWIDVA	880	1674	HIIVSGQLTDTN-INIITLFLPDDVPLNQDYH-----AKVYMTFKSPSDGTWGWGPHFVRD
Db	825	FEANSLTAEQADAMNLDANILLQASIOAHQHQLHPVPTPNAPSCWTSINTILLQWNVVA	884	1747	ETKQOQFVLINDADHDSGMTQGIIVKNIKVKYKGFNLVSIATGYAPMPDNFNSALYYWELF
Qy	881	SALHTMPSVIRTLVNIIRVYVATNALNKAESNLPQSWDEWQTLAENNEAGLSLQQOACTLADYAE	940	1729	D--KGIVTIN-----PKSILTHFESVNLNNS-----SEPMDFSGANSYFWELF



Db 1181 TGNKSGVQYETDYYELKLAHRYDGTWNTPIFDVNNKISL--KLEKNRAPGLYACG 1238  
Qy 1235 FQGBDILLVYKTKGSDPGGKNKVNAGMTIYGDGSKFQKQMENTALSRYSQLKNTDII 1294  
Db 1239 YQGBDILLVMEYNQDITLDSY--KNASMOGLIYIFADMASKDMTPEQSNVYRD--NSYQOF 1294  
Qy 1295 HTQGNDLVRKASYRPAQDFEVPASINMGS--AIGDSDITVMENGINIQTISKYSSDNLAI 1352  
Db 1295 DTNN---VRRVNNRYAEDYEIPSSVSSRKOYGMWDYILSMVYNGDIPNTYKAASSDLKI 1351  
Qy 1353 -----TLHNAFTVRDGSNGVIRNKQISAMKLTGVDGKRSQYCGNAFIANTVKKHYGGY 1405  
Db 1352 YISPKLRIIHNG-----YEGQ---KRNQCNLMNKG-----KLGRKFIV-----Y 1398  
Qy 1406 SDLG-GPIVYVNTKKNY--IASVQGHLMNADYTRLLITFPVNNYIARLFEP-PPSPNTIL 1462  
Db 1389 TSLGVNPNNSNKLMPYVYQSGNTSGLNQR--LLFHRDTTYPSPKVEAMIPGAKERSLT 1446  
Qy 1463 NTVTFTVGS-----NKTSDPKKCSYAVDGNNSQGFQIFSSYQSGWLDIDTGNNTDIK 1515  
Db 1447 NQNAAGDDYATDSLNPDKDLKQYIFMTD-----SKG-----TATDVSGPVEINTAISPAAKVQ 1499  
Qy 1516 ITVMAGSKTHFTTASDHIAFLPANSFDAMPYTFKPLEIDASSLAFTNNAIPLDIVFETKA 1575  
Db 1500 IIVKAGGKEQFTADKDVSIQSPSFDENMYQFNALEIDGSLFINNSASIDVTFAPA 1559  
Qy 1576 KQGRVLG-----KIQOTLSVKRVNNPEDIPLRETHSGAQWQVLGVTRIRLNTLLASQLV 1631  
Db 1560 EDGRKLGVESPIVTLKV-----STDNALTLHNEGAQYQWQVSYRTRLNTLTLFARQLV 1614  
Qy 1632 SRANTGIDTILMTETORLPEPLGSGFPANVLKYDPAEHGDERKFKIHNIGVGGTGR 1691  
Db 1615 ARATGIDTILSMETIQEPQLGKFVATVIPPYNLSTHGBRKFYIKHVVDNNS- 1673  
Qy 1692 QPYYSGLMSDTSSETSMILF-----VPYAEGYMHGVRVLGVGYQKITVDNTWESAFVFD 1746  
Db 1674 HIIVSGQLTDN-INIITLFIPLDDVPLNQDH-----AKVMTFKPSDGTWGWPHFVRD 1728  
Qy 1747 ETQOQFVLINDADHDSGMTQOGIVKNIKKYKGFNLVSIATGYSPAMPFNSASALYYWELF 1806  
Db 1729 D--KGIVTIN-----PKSILTHFESVNVNLNIS-----SEPMDFSGANSILYWFELF 1772  
Qy 1807 YITPMWCFORLLQEKQDEATQWVNVNYPAGYLVNGEIAPIWNCBPLETTTSWNAPL 1866  
Db 1773 YITPMLVAQRLHQBQNFDEANRWLYKYNWSPSGYIVHGQIQYQNVNRPLEDTSWNSDPL 1832  
Qy 1867 DAIDPDAVQNDPMHYKIATPMRLDQILILGDMAYRELTLDALNEAKWYVTRTLELIGD 1926  
Db 1833 DSVDPDAVAQHDPMHYKVSTPMRTLDLLIARGDHAYQLERDTLNEAKWYMQALHLLGD 1892  
Qy 1927 EPEDYGSQQWAAAPSLGSAASQTVQAAVQODITMLGRG---GVSKNLRTANSVLGLFIPY 1983  
Db 1893 KPYPFLSTWSDPLDRAADITTONAHDSAIVALRQNIPTAPLSLRASANTLTLFLPQI 1952  
Qy 1984 NPALTDYQWOTLRRLFNLRHNLSDGQPLSLAIYAETDPKALTSWQASQSGSVAVLPG 2043  
Db 1953 NEVMNMYQTLAQVYNLRHNLSDIGQPLXIPYIATPADPKALLUSAAVATVSQGGKLPES 2012  
Qy 2044 TLSLYRFPVMLERTNLVAQLTQFCTSLLSMAEHDDADLTLLILQQCMELATQSIIRIQ 2103  
Db 2013 FMSLWRFPHMLNARGVWSQLTQFGSTLQNIIEHQDAEALNALLQNAQAEILLTNSLQD 2072  
Qy 2104 RTVDEVDADIAVLAEBSRRSAQRLEKYQOQLVDEDINHGEBQAMSLDDAAQGSILAGQVLS 2163  
Db 2073 KTIELDAEKTVLEKSKAGASQSRFDSYKLYDENINAGENOAMTLRASAGLTTAVQAQR 2132  
Qy 2164 IAEGVADIVPNVGLACGSGRWGALRASASVMSLSATASQYADKISREAVERRQEW 2223  
Db 2133 LAGAADLVNPIFGFAGGSGRWGALAEATGYVMEFSANVMTEADKISQSETYRRRQEW 2192  
Qy 2224 EIQRNADGEVKQMDAQLESILKIRREAAQMOVEYQETOQAHTQAEQLLQKFTNKALYS 2283

Db 2193 EIQRNNAEALKQIDAQLKSLAVRREAAVLQKTSLKQEQTSQSLAQFLQKFSNQALYN 2252  
Qy 2284 WMRKLSAIYYQFPDLTQSFCLMAQAEALRRELTDNGVTTFIRGGAWNGTTAGLMAGETLLL 2343  
Db 2253 WLGRGLAAIYQFPDYDLAVARCLMAEQARYWELNDDSFARFKPGAWQGTYAGLLAGETJML 2312  
Qy 2344 NLAEMEKVWLERDERALERLEVRTVSLAFOYQALSSDN--FNLTEKLTQFLREGKGNVGSAG 2401  
Db 2313 SLAQMEDAHLKRDKALEVERTVSLAEVYAGLPKXNGPFSLAQEBIDKLVSQSGSAGSGN 2372  
Qy 2402 NELKL-----SNRQIEASVRLSDLKIFSDYPESIGNTRQLKQVSVTLPALVGPVEDIRAV 2456  
Db 2373 NNLAFGAGTDTKTSLOASVSFADLKIREDPASLGKIRRIKQISVTLPALILGPYQDVQAI 2432  
Qy 2457 LNYGGSIVMPRGCSAIALSHGVNDSQGMFLDNDSRYLPREGISVNDSGSITLSFPDAT- 2515  
Db 2433 LSYGDKAGLANGCEALAVSHGMNDSGQFLDNFGKLPFEGIAI--DQGTLTLSFPFNASM 2491  
Qy 2516 ---DRQKALLESLSDIILHIRYTIR 2537  
Db 2492 PEKGQATMLKTLNDIILHIRYTIK 2516

RESULT 10  
US-10-703-280-4  
; Sequence 4, Application US/10703280  
; Publication No. US2005010713A1  
; GENERAL INFORMATION:  
; APPLICANT: Dow Agrosciences LLC  
; TITLE OF INVENTION: USE OF UNTRANSLATED REGION OF OSMOTIN GENE TO ENHANCE TRANSGENE  
; TITLE OF INVENTION: EXPRESSION IN PLANTS  
; FILE REFERENCE: 61,795A  
; CURRENT APPLICATION NUMBER: US/10/703,280  
; CURRENT FILING DATE: 2003-11-07  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 2516  
; TYPE: PRT  
; ORGANISM: Photorabius luminescens  
US-10-703-280-4

Query Match 38.7%; Score 5043; DB 5; Length 2516;  
Best Local Similarity 42.6%; Pred. No. 0;  
Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

Qy 22 LADQYLSFSELKRIFDDQLSWGAEARHLXHYTHETIQKXNRLLEARIIFRANPQLSGAIRL 81  
Db 21 LTDISHSFNEFRQOVSEHLWSETHDLXHQQAQKXNRLYEARILKXNRLPQLQNAVHL 80  
Qy 82 GIERDSYR-SYDEMFGARSSSFVKGSVASMESPGAYTELYREAKDLHFESSAYHLN 140  
Db 81 AILAPNAELIYNNQFSGRASQYVAPGTVSSMFPSPAAIYTELYREARNLHASDVSYYLDT 140  
Qy 141 RRPDLADTLISQSNMDTEISTLTLSNELLLEHI--TRKTGSDSDALMESLSYRQAITPY 199  
Db 141 RRPDLKSWALSQQWMDIELSTLSNELLLESIKTESKLENYTKVWMLSTFRSGATPY 200  
Qy 200 HQPYETIRQVIMTHDSTLSALSARNPEVMQAEASILLAILANISPELYWILTEITEKNA 259  
Db 201 HDAYENREVIQDQPGLEQLNASPAIAGLHQHQAQSLGGINASISPELFNILTETEGNA 260  
Qy 260 DALFAQNFSENIIPENFASQSQSWIAKYKYLELSEVQYKYLGMQLN-GYSDSISAYVDNISTG 318  
Db 261 EELYKKNFG-NIEPASLAMPEYLYKRYNLSDEELUSQFIGNKASNFGQOEYSN---NQLITP 316  
Qy 319 LVVNESKLEAYKITR-VKTDVDDKINIFYDLMEYGNQFPFRANFKVSRFEGAT----- 372  
Db 317 VNNSDGTVKYRITREYTTNAYQMDVELFP--FGGEN---YRLDYKFKFNYSYLSIK 371  
Qy 373 -----LRKNAGPSGIVSGPLIANTWTFKSNYLSNISDSEYKNGVKIYAYRYTSTS 425  
Db 372 LNDKRELVRTEGAPO-----VNIEYSANITLNTAD--ISQPFELGLTRVLPSSG 418















Qy 1286 QLKNTEDIHQND-----LVKASVFPAODEEVPASL--NMGSALGDDSLTVMXGNI 1338  
 Db 1271 NSYQFDTVMADPDSNDKKVITRRVNNRYAEDIEPSSVTSNSYMWGDHSLTWMYGGSV 1330  
 Qy 1339 POITSKYSSDLATITLHNAAFTRVYDGSNIRNKQISAMKLTGVDGKSGYGNFIANT 1398  
 Db 1331 PNITFESAEDRLST-NWALSIIHNGYAGT-RRIQCNLMK----- 1369  
 Qy 1399 VKHYGGYSDLGPIITVYNNKTKRYIASVQGHLMNADYTRRLILTPVNNYARLPEFPSP 1458  
 Db 1370 -----QYASLGDKFTIYDS-----SFDANRPNLVP-----LPKFGKDE 1403  
 Qy 1459 NT-----ILNT-----VFTVGSNKTDFPKCSYAVD-GNNSQGFQ----- 1492  
 Db 1404 NSDDSICTYNNPSSSEDKWTFSSKDDNKTDYNGGTQCIDAGTSGDYNNLQIEVIS 1463  
 Qy 1493 ----IFSSYQSGWLDIDTGINNTDIIKITVMAGSKTHFTTA--SDHIASLPANSFDDAMPY 1546  
 Db 1464 VTGGYWSYKISNPININTGDSAKVYTVKAGGDDQIFTDADNTYVPPQAPSPFEEMIY 1523  
 Qy 1547 TFKPLEIDASSIAFTNNIAPIDIVFETKAKGRVLGKIQTLVSKRVNPNPEDILFRET 1606  
 Db 1524 QFNNTIDCKNLNFIDNQAHEIDFTATAQDGRFLGAETFIIPVTKVLGTENVIALYSE 1583  
 Qy 1607 HSGAQYMLGVYRIRLNTLLASQLSVSRANTGIDTILMTETORLPEPPLGEGFEANFVLPK 1666  
 Db 1584 NNGVQYMOIGAYRRLNTLPAQLVSRANRGIDAVLSMETQNIQBPQLGAGTYVQLVLDK 1643  
 Qy 1667 YDPAEHGDERWFKIHIHGVNGTGRQPYISGMLDTSMTLFPVY---ABGYMHEGV 1723  
 Db 1644 YDESHTGNTKSAIEYVDIIPRENDSEFVYIQGELSETSQTVVKVFLSYFIATGKNHLWV 1703  
 Qy 1724 RLGVGYKITYDNTWESAFFYDETKQO-----FVLINDADHDGWTQOQIVGNIKYKGF 1779  
 Db 1704 R--AKYQKETDK-----ILFDRDDEKDPHGWFSLDDHKTFSGLSSAQLKN----- 1748  
 Qy 1780 LNVSTATGYSAPMDPNSASALYYELFYTPMCMCFQRLLOSKQFDEATQWNNVYNPAGY 1839  
 Db 1749 -----DSEPMDFSGANALFYELFYTPMMAHRLLOEQNFDAANHWFRYVWSPGY 1800  
 Qy 1840 IVNGEIAPIWNCPRLETTSWNANPLDAIDPDAVAQDPMHYKIATPMRLDQLILRGD 1899  
 Db 1801 IVDGKIALYHNVNPRELEEDTSWNAQQLDSTDPDAVAQDDPMHYKVATPMATLIDLMLARGD 1860  
 Qy 1900 MAYRELTDALNEAKWVTRLELIGDEPBYGSOQWAAPSLSGAASQTQAAQOQDLTM 1959  
 Db 1861 AAYROLERDTLAEAKWVYTOALNLLIGDEPVMVLTSTWANPTLIGNAASKTTQOVRQOVLTO 1920  
 Qy 1960 LGRGVSKN--LRTANSIVGLFLPEYNPALTDYQTLRLRLNLRHNLSDIGQPLSLAIY 2017  
 Db 1921 LRLNSRVKTPLLGTANSUTALFLPOENSKLKGWYRTLAQRNPNLNRHNLSDIGQPLSLPLY 1980  
 Qy 2018 AEPTDPKALLTSMVQASQGSNAVLPGTLLSYRFPVMLERTNLNAQLTQFGTSLLSMABH 2077  
 Db 1981 AKPADPKALLSNAVSASQGGADLPKAPLTIHRFPQMLEGARGLVNQLLQFGSSLIGYSER 2040  
 Qy 2078 DDADELTLTLCOQHELTATQIRIQORTVDEVDADIAVLESRSQAQRLEKYQOYLDED 2137  
 Db 2041 QDAEASQLLOQASELITSTIRMQDNLQALDESEKTLQVSLAGVQQRFDYSQSLYEEN 2100  
 Qy 2138 INHGEORAMSLDDRAAGOSLAGOVLSIAEGVADIVPNVFGACGSRGWAALRASAVMS 2197  
 Db 2101 INAGEORALARSESAIESQGAQISRMAGAGVDMAPNIFGLADGGMHYGATAYAIADGIE 2160  
 Qy 2198 LSATASOYSAKISIRSEYRRRRQEWBIORDNADGEVQMDAQLESKIRREAAQOMVEY 2257  
 Db 2161 LSASAKVYDAEKVAQSEIYRRRRQEWKIQORDNAQAEINQLNAQLESLSIRREAAEMQKEY 2220  
 Qy 2258 QETQOATQAOQLLELQRFETKALYSWRGKLSAIYQFFDLTQSFCLMAQEARLRELT 2317  
 Db 2221 LKTOQAQAQALTFURSKPSNALYSWLRGLSIGIYFQFDLAVSRCLMAQESQYOWEAND 2280  
 Qy 2318 NGVTFIRGAGWNGTTAGLMAGETLLNLNAEMEKWLERDERALBVRTVSLAQFYQAL-S 2376

Db 2281 NSISFVPGAWQGTAGLCCGEALIQNLAQMEAYLKWESRALEVERTVSLAVVYDSLEG 2340  
 Qy 2377 SDNPNLFEKUTQFIRBEKGNVAGSNELKLSNRQIEASVRLSDLKIFSDYPESL---GNT 2433  
 Db 2341 NDRNLAEQIPALLDKGEGTAGTKENGSLANATLSASVKLSDLKLGATDYDPSIVGSKV 2400  
 Qy 2434 ROLKOVSVTIPALVGPVEDIRAVLNYGSGIVMPRGCSAIALSHGVNDSQFQMLDFNDSRY 2493  
 Db 2401 RRIKQISVSPALVGPYQDVQAMLSYGGSTQLPKGCASALAVSHGTNDSGQQLDFNDGKY 2460  
 Qy 2494 LPFEGISVNDSGSLTSLFFPDATDRQKALLBSLSLIIILHRYTIR 2537  
 Db 2461 LPFEGIALDDQGTNLNLFQPNATQKQAILQTMDSIILHRYTIR 2504

RESULT 14  
 US-11-020-848-11  
 ; Sequence 11, Application US/11020848  
 ; Publication No. US20050155104A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Apel-Birkhold, Patricia  
 ; APPLICANT: Hey, Timothy  
 ; APPLICANT: Sheets, Joel  
 ; APPLICANT: Meade, Tom  
 ; APPLICANT: Li, Ze Sheng  
 ; APPLICANT: Lira, Justin  
 ; APPLICANT: Russell, Sean  
 ; APPLICANT: Thompson, Robin  
 ; APPLICANT: Mitchell, Jon  
 ; APPLICANT: Fencil, Kristin  
 ; TITLE OF INVENTION: Toxin Complex Proteins and Genes from Xenorhabdus bovienii  
 ; FILE REFERENCE: DAS-114XCL  
 ; CURRENT APPLICATION NUMBER: US/11/020,848  
 ; CURRENT FILING DATE: 2004-12-23  
 ; PRIOR APPLICATION NUMBER: US 60/534,893  
 ; PRIOR FILING DATE: 2004-01-07  
 ; NUMBER OF SEQ ID NOS: 11  
 ; SOFTWARE: Patent in version 3.2  
 ; SEQ ID NO 11  
 ; LENGTH: 2534  
 ; TYPE: PRT  
 ; ORGANISM: Xenorhabdus bovienii  
 US-11-020-848-11

Query Match 38.2%; Score 4976; DB 6; Length 2534;  
 Best Local Similarity 41.4%; Pred. No. 0;  
 Matches 1089; Conservative 490; Mismatches 861; Indels 190; Gaps 49;

Qy 1 MYSTAVLLNKISPTRDGQMTLADLQVLSFSELRKIFDDQLSWGAEARHLYHETIEQKNN 60  
 Db 1 MINTENILIRUNRESSQEPMTLAHIMFISFAFRKEVKDTLWNGESHLYLAAKAEKEN 60  
 Qy 61 RLLEARIPTRANPOLSGAIRLGIERSYSRSYDEMFGARSSSVFKPGSVASMFSPAGYLT 120  
 Db 61 RIFEARULLSRANPOLRGAVRIGIQQLSQRSQSYDTLFGGRSGKYLPGSVASMFSPAYLT 120  
 Qy 121 ELYEAKDLHPSSAYHLNDRPDLADLTLSQSNMDTEISTLTLSNELLEHI-TRKGG 179  
 Db 121 ELYRESRHLHSESIYHLDKRPPDLQSLIMLTQENQDQTLSTLSLNSDLTFDGIKNKKLN 180  
 Qy 180 DSDALMESLSYRQAITDPYHOPVETIROVIMTHDSTLSLSRNPENVNGQAEASLLAIL 239  
 Db 181 KNEDVLKMLSDWRLSGNTPYHQPETLSNIVSQDLPQLSQVSPKVLGLLSVPSLLGIS 240  
 Qy 240 ANISPELYNLTETITEKNADALFAQNFSENIPTENFASQSWIAKYIGLEISEVQKYLGM 299  
 Db 241 SQISPELYKILTEBITAENAQDMYKKNFGD-LPISALSNPNLYMLMKYDIDADTLRAVMGI 299  
 Qy 300 LQNGYSSTSAVVDNISGLVNNESKLEAYKITRVKTDYDDKNNYFDLAVYEGNNOFFI 359  
 Db 300 YGSGQNDDEPAFISDQAIVTYVLLDDKNSFVTLITRTKETGYDMQVNFTEIPTKDGKUY 359



Db 2424 AARRIKQVSVSLPALLGPVQVQAILSYSDMKGIPKGCASALAVSNGMNDSSQFQOLDPND 2483

QY 2491 SRYLPFEGISV---NDSSGLTSLPDPDATDRQKALLLESIDILHRYTIR 2537

Db 2484 TKYLPFEGINIPKDKQDQALVSPFNADAKQKTMLLSLSDILHRYTIR 2533

RESULT 15

US-10-706-424-8

; Sequence 8, Application US/10706424

; Publication No. US20040103455A1

; GENERAL INFORMATION:

; APPLICANT: ffrrench-Constant, Richard

; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens

; FILE REFERENCE: 62878

; CURRENT APPLICATION NUMBER: US/10/706,424

; CURRENT FILING DATE: 2003-11-12

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 8

; LENGTH: 2381

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-10-706-424-8

Query Match 36.8%; Score 4802; DB 4; Length 2381;

Best Local Similarity 40.8%; Pred. No. 0;

Matches 1053; Conservative 449; Mismatches 793; Indels 284; Gaps 48;

QY 22 LADQLYSFSELRKIFDDOLSGEARHLYHTEIOKKNRLLEARIFTRANPOLSGAIRL 81

Db 21 LTDICHYSFNRQVQSDDLSEWSETRNRYRQAEQENQLYEARILKRANPOLQNAVHL 80

QY 82 GIERDSVS-RSYDEMFGARSSSFVKPGSVASMFSPAGYLTYELTYREAKDLHFSSSAYHLDN 140

Db 81 GITLPHAEELRGVSEFGGRASQYVAPGVSVMFSPAAYLTLYREARNLHSDSVYHLDE 140

QY 141 RRPDLADTLQSNDMTISITLSNELLLEHI-TRKTCGGSDA-LMESLSTYQRAIDTP 198

Db 141 RRPDLQSMTLQSQNDMTLSLSNEILLKGIKANQSNLSDTKWMEMLSTFPSPGTIP 200

QY 199 YHQPVTETROVIMTHDTSLSALSRNPEYMGQAGSLLAILANISPELYNLTITEITKN 258

Db 201 YHDAYENVRKAIQDPKLEQFQKSPAVAGLHQASLLGINNISPELNFILTEITKAN 260

QY 259 ADALFAQNFSENITPENFASQSWIAKYGLSELSEVQKGLQMLQNGYSDSTSAYVDNISTG 318

Db 261 AEATYKQFQD-IDPACLAMPEYLSKSYNFSDELSQFI-----RKYPDN----- 304

QY 319 LVVNNESKLEAYKITRVKTDYDKNINYPDLMEGNNQFFIRANPKVSRFEGATLRKNAG 378

Db 305 -----ELNTO----- 309

QY 379 PSGIVGSLGPLIANTNPKSNVLSNISDSEYKNGVKIVAYRVSTSTSATNQGQGIPTES 438

Db 310 ----- 309

QY 439 YPLTIFALKKRAIRLCHLTSGLSPELQTIIVRSNDAQIINDSVLTQVFTLYFSHYAL 498

Db 310 ---KHLKLKINKILLQSQVNLPLKLDIIEIQN-----ITPTVLGKFLVLYKVMQYNI 362

QY 499 SFDDAQLVNGSVINQYADDDSVSHNRLFNTPPLKGI FEADGNVTSIDPBEQSTFARS 558

Db 363 GTETALILCNDSISQYSYNSQPSQDRLFNTSPLNGQYFVIEDNTIDLSLNSTDN-WHKA 421

QY 559 ALMRGLGVNSGELYQGLGVLDAQNTITLSVFVSSLYRLTLTLARVHQLTVNELCML- 617

Db 422 VLKRAFNVDDLSLRLLHIANHNTDGIANNKMLSNLYMTKLADLHQITIDELYLL 481

QY 618 ---YGLSPFNGKGTASLSSGELPRLVIMLYQVQWMLTEABITTEAIWLCTPEFSGNISPE 675

Db 482 ITIGEDKIN---LVDIDDKELEKILNRLOTLSNMLHTQKWSIYQLFLMTTNYDKLTPTPE 538

QY 676 ISNLNLRPSISEDMAQSHNRELQAEILADPFIATLHLASPDMARYILLWTDNLRPGGL 735

Db 539 IONLLDTVYNGL-QNFKNKTKLLAA--IAYIATLQLPSENVAHSILLWADKIKPSEN 595

QY 736 DIAGFMTLVLKESLNANETT-----QLVQFCHVMAQSLSVQTLRLSEAEASLVLSVG 788

Db 596 KITAEKFWLQNRDTELSKPPMQEQIIOYCHCLAQLTWIYRSSGINENAFRLFIEKP 655

QY 789 --FAVLGAKNQAPAGOHNIDTLFSLYRPHOWINGLGNPGSDTLDMLRQOTLTADRLASVWG 846

Db 656 TIFGIPDEPNKATPAHNAPTLIILTRFANWVNSGKASPIITAFENKTLTAEKLANAMN 715

QY 847 LDISWVTOAMVSA-----GVNQLQWODINTVLQWIDVASALHTMPSVIRTLVNIYV 899

Db 716 LDANLLQASIQAOANYQKVYKENTFSNWSQSDIILQWNTINASNLNISQGISPLIALDYI 775

QY 900 TALNKAESNLPSEDWOTLAENMEAGLSTQAOQTLADYTABERLSSVLCNFWLANIOPRGV 959

Db 776 ---KPAQKTPTYAQWENAAIALTAGLDTQOQTHLHVFLDESRTALSNYYIGKVANRAA 831

QY 960 SLHSRDDLYSYFLIDNQVSSAIKTRRLAEATAGIOLYINRALNRIEPNARADVSTROFFT 1019

Db 832 SIKSRDLYQYLLIDNQVSABIKTRIAEATIASIQLYVNRALENIETHAVSDVITROFFI 891

QY 1020 DW-TVNNRYSTWGVSVRLVYYPENIDPTORIGOTRMWDELLENISQSKLSDTVEDAFK 1078

Db 892 DWDKYNKRYSTWAGVSQLVYYPENIDPTMRIGOTKMMDTLQSVQSQQLNADTVEDAFK 951

QY 1079 TYLRFETVADLVKVSAYHDNVNSNTGLTWFGVGTRENLPEYYWRNVDSISRMAGELAA 1138

Db 952 SYLTSFEQVANLEVISAYHDNVNDQGLTYPIGSKTEVNYWRSVDHSKFNDGKFAAN 1011

QY 1139 AWKQWTIKDTAVNPYKDAIRPVI FRERHLHIWEKEEVAKNGTDPVETD-RFTILKLAFL 1197

Db 1012 AWSEWHKIDCAINPYQSTIRPVIYKSLYLIWLQKETAKQEDNKVTTDYHELKLAHI 1071

QY 1198 RHDCGWSAPWSYDITTOVEAVTDKPDTERLALAASGFQGBDTELLVYVYKTKGYSDFGG 1257

Db 1072 RYDGTWNVPIITFDVEKILALELYK--SQAPGLCAGYQGEDTELLIIMFYRKKEKLDDY-- 1127

QY 1258 SNKXVAGMTIYDGSFKKMENTALSRYSQLKXNTFDIHTQGNDLVRKASYRFAQDFEVA 1317

Db 1128 KTAQMOCFYIFSDMSSKDMTNEQCNSYRDNGYTHFTDNTSDNSVIR-INNRVYAEDEYEPS 1186

QY 1318 SLNMGSA--IGDSSLTWENGNIPQITSKYSSDNLAITLHNAAFTRYDGSNVRNKQI 1375

Db 1187 LINSNSHDWGEYNLSQVYGGNI-VINYKVTSDNLKIYI----- 1224

QY 1376 SAMKLTGV-DGKSQVGNAPFIANTVKHYGG-----YSDLG-----GPITVY 1415

Db 1225 -SPKLRIIHDKQE--GREIQSNIKKYKGLGKDFIYTSLGINPNSSNPFMPFVYQY 1281

QY 1416 NKTKNYIASVQGHLMNADYTRRLITPVENNYARLFEF-PFSPNTILNTVTVGS----- 1470

Db 1282 NGNTSGLA--QGRLL---FHR-----DTSYSKVAAWIPGAGKSLINENANIGDDCAE 1329

QY 1471 ---NKTSDFKKCSYAVDGNNSQGFQIFSSYQSSGWLIDITDGINNTDILKITWAGSKTHFF 1527

Db 1330 DSVNKPDDLQYIYMTD---SKG---TATDVSQFVDINTAISSEKVOITTKAG-KEYSL 1381

QY 1528 TASDHIASLPANSPDAMPYTPKPLEIDASSIAFTNNIAPLDIVFETKADGRVLG-KIKQ 1586

Db 1382 TANKDVSQPSFSEEMCYQFNALIEDGSNLFNTNNSIDVTTALADDERKGLGYELFN 1441

QY 1587 TLSVKRYNPNEDILFLRETHSGAQYMQGLGYVIRLANTLLASQLVSRANTGIDTILTWT 1646

Db 1442 IPVLIQV--KTDNALTLFHDENGAQYMCWQAYRIRLNTLFAQLVERANTGIDTILSMET 1499

QY 1647 QRLPEPPLGEGFFANFVLPKYDPAEHGDERWFKTHIGNVGNTGROPYVSGMLSDTSETS 1706

Db 1500 QNIOEPMWGIGAYIELIDKYNPDIGHNTGSKFIYIGDIFKAGDHFPIYQOALSIDITQTT 1559



**THIS PAGE BLANK (USPTO)**

---



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	5043	38.7	2516	6	US-10-647-956A-2	Sequence 2, Appli
2	4986.5	38.2	2504	6	US-10-647-956A-8	Sequence 8, Appli
3	246	1.9	246	6	US-10-793-626-2964	Sequence 2964, Ap
4	222	1.7	2710	7	US-11-051-453-41	Sequence 41, Appl
5	222	1.7	5291	7	US-11-052-554A-281	Sequence 281, App
6	221.5	1.7	2902	7	US-11-052-554A-91	Sequence 91, Appl
7	202.5	1.6	3194	7	US-11-052-554A-90	Sequence 90, Appl
8	201.5	1.5	2367	7	US-11-051-453-42	Sequence 42, Appl
9	192	1.5	5935	6	US-10-995-561-776	Sequence 776, App
10	191.5	1.5	2399	7	US-11-052-554A-92	Sequence 92, Appl
11	182	1.4	2053	7	US-11-013-759-9	Sequence 9, Appli
12	178.5	1.4	1889	7	US-11-102-476-46	Sequence 46, Appl
13	170.5	1.3	2087	7	US-11-075-185-28	Sequence 28, Appl
14	167.5	1.3	2769	7	US-11-113-424-14	Sequence 14, Appl
15	165	1.3	2340	7	US-11-052-554A-171	Sequence 171, App
16	160	1.2	1461	7	US-11-052-554A-283	Sequence 283, App
17	160	1.2	1360	7	US-11-069-834-50	Sequence 50, Appl
18	158.5	1.2	1242	7	US-11-019-711-110	Sequence 110, App
19	157	1.2	1382	7	US-11-098-686-10934	Sequence 10934, A
20	155	1.2	980	7	US-11-052-554A-17	Sequence 17, Appl
21	155	1.2	2668	7	US-11-124-368A-215	Sequence 215, App
22	153.5	1.2	1296	6	US-10-615-668-3	Sequence 3, Appli
23	153.5	1.2	1449	7	US-11-052-554A-237	Sequence 237, App
24	152.5	1.2	2015	7	US-11-052-554A-374	Sequence 374, App
25	152	1.2	5406	6	US-10-995-561-774	Sequence 774, App



Db	2373	NNLAFGAGTDTKTSLQASVSFADUKTREDYPASIGKTRIRKQISVTLPALGYPQDVQAI	2432
Qy	2457	LNYGGSIVMPRGCSAIALSHGVNDSGQFMLDVNDSPLEPGISVNDSGSLTILSPFDDAT	2515
Db	2433	LSYGDRAKLANGCEALAVSHGVNDSGQFQDQFNDGKFLPEGIAI	2491
Qy	2516	---DRQKALLIESDIILHRYTIR	2537
Db	2492	PEKQKATMLKTLNDIILHRYTIK	2516

```

RESULT 2
US-10-647-956A-8
; Sequence 8, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/10/647,956A
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-647-956A-8

```

[illegible]

Qy	473	NAGIINDSVLTKVFTYTLFYSHRYALSFDDAQVNGSVINQYADDDSVSHFNRLFWNTPL	532
Db	469	NSTKSIITVELNKVYRKFYIDRYGISEETAAILANINISQQAVGNQLOFQELFNHPPL	528
Qy	533	KGKIFE-ADGNT-----VSDPDEEQSTFARSALMRGLGVNSGELYQL-----GKLAG	579
Db	529	NGRIYELISDUNSXLNPNPDNLKLPDSTGDDORCAVLKRAPOVNASELYQMLLITDRKEDG	588
Qy	580	VL--DAQNTITLSVFISSLYRLTLARVHQLTVNELCMLYGLSPFNGKTTASLGGELP	637
Db	589	VIKNNLEN-----LSDLYVSLLAQIAHNLIAELNILLVLCGYGDTWYIQTIDNLA	640
Qy	638	RLVWLKYQWNTBEAITTEALTWLCTBFGSNISPEISNL-----LANNLRPISED	690
Db	641	KIVETLLWITQWLTKQWTTDLFLMTTATYSTLTPTESNLTATLSTLHGKESLIGED	700
Qy	691	MAOSHNRLEQAEILAPFIATLHSLASPDMARYILLMTDNLRLPCGLDIAGFMTLVLKESLN	750
Db	701	LKRA-----MACPFTSALHLTQOEVAYDULLLWIDQIQPAQIITVDGEW-----EEVQ	746
Qy	751	ANETT-QLVQFCHVMAQLSLSVQTLSELAELSVLVISPAVLGAKNQAPAGQHNDITLFS	809
Db	747	TTPTSLKVTFAQVLAQLSLIYRRIGLSETELSILVTQSSLLVAGKS--ILDHGLLTLMA	804
Qy	810	LYRFHOWINGLNGPSSDTLMDLROQTILTADRSLASVMGLDISWYTOA--MYSAVGNQLQCV	867
Db	805	LEGFHTWVNGLQHASLILAAALKDGAULTVDVAQANMKESLQLQMAANOKEKOLTKLTSW	864
Qy	868	QDINTVLQIDVASALHTWPSVIRTLVNIYRYTALNKABSNLPSWDEWOTLAENMEAGLS	927
Db	865	TQIDAILQWLQMSALAVSPLDIAGMALKYIGIDHNYA-----AMQAAALMAD-H	915
Qy	928	TQOQATLADYTBRLSSVLNCWFLANIQPEGVSLHSRDDLISYFLIDNQVSSAIKTRRLA	987
Db	916	ANQAQKLD---ETFSKALCNYINAVDSAAGVDRDNGLYTYLLIDNQVSAVDITSRIA	972
Qy	988	EAIAGIQLYINRALNRIEPNARADVSTROFTDW--TVNNRYSTWGVSRVLVYPENYIDP	1046
Db	973	EAIAGIQLYVNRALNRDEGLASDVSTROFTDWERYNKRYSTWAGVSELVYYPENYVDP	1032
Qy	1047	TORIGOTRWDELLENISQSLSRDVTEDAFKTYLTFETVADLKVVSAYHDNVSNTGL	1106
Db	1033	TORIGOTKMDALLQINSQLMADTVEDAFKTYLTSFEQVANKLVISAYHDNWNVDQGL	1092
Qy	1107	TWFGQOTRENLPYYWRNVDISRMOAGELAAANKEWTKIDTAVNPVKDAIRPVIFRERL	1166
Db	1093	TYFIGIDQAPGTYVYRSVDHSCENGKFPANAAGWENKITCAVNPWKNIIRPVYMSRL	1152
Qy	1167	HLIWBKBEVAKNGTDPVETYDRFTLKLAPLRHGDGWSAPWSYDITQVEAVTDXKPDTE	1226
Db	1153	YLLWLBEQ--SKKSDGKTTIYQYNLKLARIHVYDGSWNTPFTDVTKEKVNYTSS7DAAE	1210
Qy	1227	RLAALASGQGEDTILVFYTKGKSYSDGGSKNVAGMTIYDGSFKKMENTALSRY-S	1285
Db	1211	SLGLYCTGYQGEDTLLVMFYMQSSYSSYTDNNAPVTGLYIFADMSDDNNTNAQAITYWN	1270
Qy	1286	QLKNTFDIHTQOND---LVKASGYRAQDPEVPASL--NNGSAIGDDSLTWMENGNI	1338
Db	1271	NSVPQEDTWADPDSDNKKVITRVNVRNRYAEDYEIPESSVTSNSNYSGWDSHLTMYCGSV	1330
Qy	1339	POITSKYSSDNLAITLHNAAFVRYDGSNGVIRNKQISAMKULGVDPCKSGOYGNAFIANT	1398
Db	1331	PNITFESAEDLRST-NMALSIHNGYAGT--RRIQCNLMK-----	1369
Qy	1399	VKHYGYSIDLGGPIYTNKTKNYIASVOGHLMNADYTRRLILTVPENNYIARLFEPPFSP	1458
Db	1370	-----QYASLGDKFIIYDS-----SFDANRFLVLP-----LFKFGKDE	1403
Qy	1459	NT-----ILNT-----VFTVGSNKTSDPKKCSYVD--GNNSQGPQ-----	1492
Db	1404	NSDSTICINENPSSBKKWYFSSKODNKTAYNGTQCIADGTSKDNKFTYNNQIEIVIS	1463
Qy	1493	-----IFSSYQSSGLWIDTGINNTDITKITVMAGSKTHTFTA--SDRIASLPANSFPAMPY	1546

```

Db      1464 VTGGWSSYKISNININTGSDSAKVVVKVAGGDDQIFADNNTYVPPQAPSEEMII 1523
Qy      1547 TFKPLEIDASSLATNNIADIDIVPEYKAKDGRVLGKIKQTLVKRVNYPNEDILFIRET 1606
Db      1524 QFNNTLIDCKNLNFDINQAHEIDFTATAQDGRFLGAETFIPTVKVGLGTENNIALYSE 1583
Qy      1607 HSGQYVQMGVYVRLTALNTLASQVLSRANTGIDITLMTETORLPEPPLGEGFFANFVLPK 1666
Db      1584 NNGVYMQIGAYRRLNTLFAQQVLSRANRIGIDAVLSMETONIPEPGLAGTYVQVLVDK 1643
Qy      1667 YDPAEHGDERFKIHIGNVGNTGQPYSGMLSDTSETMTLFPVY---AEGYMHGV 1723
Db      1644 YDESIHGNTKSFALYVDIIPKENSFVLYQELSETSGTVVKVFLSFIEATGKNHLW 1703
Qy      1724 RLGVGYKITYDNTWESAFFYDETKQO-----FVLINDADHSGMTQOGIVKNIKKYKGF 1779
Db      1704 R--AKYQKETDK-----ILFDRDTEKDPHGWFSLDDHKTFSGLSSAQAALKN----- 1748
Qy      1780 LNVSIATGYSAPMDPNSASALYYWELFVYTPMCMCFORLLQEKQDEATOMINYVNPAGY 1839
Db      1749 -----DSEPMDFSGANALFYWELFYTPMMWAHRLQLQONFDAQNHFRYVWSPSGY 1800
Qy      1840 IVNGEIAPIWNCRLPSETTSWNNANPLDAIDPDAVQNDPMHYKIATFMRLDOLILRGD 1899
Db      1801 IVDGKIALYHWNVPLEEDTSSWNAQQLDSDTPDAVAQDDPMHYKVAIFMATLILLMARGD 1860
Qy      1900 MAYRELTRDALNEAKMYVTRTLELLGDPEPDYGSQWAAAPSLSGAASQTVOAAVQODLTM 1959
Db      1861 AAYRQLERDITLAEAKMYVTOALNLLGDPEVMLSTTWANPTLGNAAKTTQOVRQOVLQT 1920
Qy      1960 LGRGVSKN--LRTANSIAGVLFPEYNPALTDYQTLRLFLNRLNLSIDGQPLSLAIY 2017
Db      1921 LRLNSRVKTPPLGANSUTALFLPOENSCLKGYWRTLAQRMFNLRHLSIDGQPLSLPLY 1980
Qy      2018 AEPTDPKALLTSMVQASQGSAAVLPGLTSLYRFPVMLERTNLVAQLTQFOTSLLSMAEH 2077
Db      1981 AKPADPKALLSAAVSASQGGADLPKAPLTIHRFPQMLEGAGLVNQLIQFGSSLLGYSER 2040
Qy      2078 DDADELTLTLLQQOMELATQSIIRIQORTVDEVDADIIVLAERSSAQNREKYQOVLDED 2137
Db      2041 QDAEAMSQLLQTAQASELITLSIRMQDNLQALDESEKTAQLVSLAGVQORFDSYSLYEEN 2100
Qy      2138 INHGEORAMSLDDAAAGSLAGOVLSAEGVADIVPNVFGIACGSRWGAALRASAVMS 2197
Db      2101 INAGEORALARSSESATESQQAQISRWAGAGVDWAPNIFGLADGGMHYGAIYAIADGIE 2160
Qy      2198 LSATASQVSADKISRSEAYRRRQEWETIQRDNADGEVKQMDAQLESKIRREAAQMVEY 2257
Db      2161 LSASAKWDAEKVAQSEIYRRRQEWKIQRDNAQAEINQLNAQLESLSIRREAAEMQKEY 2220
Qy      2258 QETQOAHQAOLEQLLQKFTNKALYSWMRGKLSIYYQFPDLTOSFCLMAQEARLRELT 2317
Db      2221 LKTOQAQAQOALTFLRSKFSNQALYSWLRGLSGIYFQFYDLAVSRCLMAEQSQWEAND 2280
Qy      2318 NGVTFIRGGWNGTGTAGIMAGETILLNLAEMEKVWLERDERALKVTRTVSLAQFYQAL-S 2376
Db      2281 NSISFVPGWQGYTAGLUGCEALIQNALQWEEAYLKWESRALVERTVSLAVYDSLEG 2340
Qy      2377 SDNFNLTKEKTLQFLREGKNGVSGNELKLSNRQIEASVRLSDLKIFSDYPESL---GNT 2433
Db      2341 NDRFNLAEQIPALLDKGEGTAGTKENGISLANAILLSASVKLSDLKLGTDYDPSIVGSKV 2400
Qy      2434 ROLKQVSTLPAIVGPVEDIRAVLNYGGSIVWPGCSAIALSHGVNDSGQPMFLDPNDSRY 2493
Db      2401 RRIKQISVSLPALVGPVQDVQAMLSYGGSTQLPKGCSALAVSHGTNDSGQQLDFNDGKY 2460
Qy      2494 LPFEGISVNDSSGSLTLPSPDATDROKALLESLSDIILHIRVTIR 2537
Db      2461 LPFEGIALDDQGTNLNLPNPATDKQKAILQTMDSIILHIRVTIR 2504

```













Db 1297 NSSNLSSGLYQMOAKSVLPDMS-----NLSVSVGTSSIKAINLSQNASINA 1344  
Qy 171 -EHITRKGGD---SDALMESLSTYRQADITPYHQPYETIROVIMTHDSTLS---ALGRN 223  
Db 1345 SNHSTLELQGLDNLNDT---SSLNLQSAINVSNNATINDYASLIASNGSHLNFCAVNFN 1402  
Qy 224 PEVMQAGASLAILANISPELYNLTETETKRNADLFAQ-NFSEN-----IT 272  
Db 1403 S-----ANITTSL---SSSIVFKGAVSLRGOFNLNNSSLDFQGSASIT 1444  
Qy 273 P-----ENPASQSWIAKYCYLELSEVQKYLGMONGYSOSTSAYVDNISITGLVNNES 325  
Db 1445 SNTAFNFYDNFASQSPITFHQALDIKVLPLSGNLN---PNSSVNLNKNLSQLVFSQOG 1501  
Qy 326 KLEAVKIRVKTDDYDKNINFDLMVEGNNOPIRANFKVSREFCATLRKAG----- 378  
Db 1502 SLNIANIDLLSLNGKRVYNI IQADMNGWYERINP-----FG-WRINDGIYDAKQ 1554  
Qy 379 PSGIVGSLGPIANTFNKSNYLS-----NISSEYKNGVKIYAYRYTSTSTATNOGG 431  
Db 1555 TVSFNPLNNALKITESPKNNQLSVTLSPQIKNTLYNISGEIFNYQ-----KVYNNAN 1609  
Qy 432 GIFTPESEPLTIFALKINKAKRLCLTSLSPNELQTIIVRSNAQGIINDSVLTKVPTLP 491  
Db 1610 GYISY-----SDDAQGVYLTSSVKGY--- 1632  
Qy 492 YSHRYALSFDDAQVLNGSVINQYADDDSVSHPNRLFTPLKGIKFEADGNTVSDPDEE 551  
Db 1633 -----NPNQSYQASGNTTKNNLTSESSVISTQYNAQNPIS----- 1671  
Qy 552 OSTPARSALMRGLVNGSELVOLGKLAGVLDAQ-NTITLSVFPVTSLSYRLTLARVHOLT 610  
Db 1672 -----ALHVVNGYFNFSNIKALQWALKLYPEIKKILGNDFSLSSNLKAGDA-LNOLT 1724  
Qy 611 -----VNELCMLYCLSPFNCKTTASLSSGELPRIVLWLYQVOWLFEABITTEAI 660  
Db 1725 KLITPSDMKNINEL-----IDNANNSVVQNFNG---TLIIGATKIGQTDNLSAVVFG 1776  
Qy 661 W-----LICTPEFSNISPEISNLNINLRPSISDMAQSHNRELQABILAPFAAT 711  
Db 1777 GYQKPCDYDVC-QKPRG-----TYLGQLLESISADLGVIDTTFNAKEI---YLTGT 1825  
Qy 712 LHLASP-----DMARYILWTDNLRPGGLDIAGFTWLVLKESLNANETQLVQF 760  
Db 1826 LSGSNAGWTGGSASVTFNSQTSILNQAIVSSQTD--GIFSMLEQEGI--NKVFNQAGL 1881  
Qy 761 CHVMAQL-----SLSVQTLRLSEAELSVLVSGFAVLGAKNOP-----AGOHNI 804  
Db 1882 ANILGEVAMQSINKAGGLGNLTIVNTLG-----SDSVIGGYLTPEQKNQTLSQLLGQNNF 1935  
Qy 805 DTLPSLYRFHOWINGLGNPGSDTL--DMLRQO---TLTADRLASVMGLDISMVTQAMVS 858  
Db 1936 DNL-----MNSGLNTAKDLIRQKLGFTWGLVGLAGLGIDQNPKEKLG 1983  
Qy 859 AGVNOLOCQDIT-NTVLQWIDVASALHTMPSVIRTLNIRVYVYALNKAESNLPWDEWOT 917  
Db 1984 MSINDLLSKGLFNQITGFISANDIGQIVSMQDIVK-----PS-----D 2024  
Qy 918 LAENMEAGLSTQO-AQTLADYTAERLSSVLGFWPLANTQPECVSLHSRDDLYSYFLIDNQ 976  
Db 2025 ALKNDVAALGQKMGIEFLGQDTFNSLSLLQN-----OQ 2058  
Qy 977 VSSAKTTLRLAIAIGIOLYNRLNRLNTEPN-----ARADVSTRQFFTD 1020  
Db 2059 IKSVDKVLAKGLSI---YEQGLGDLIPNLKKGIFAPYGLSQVQKGFSPNAQGVN 2115  
Qy 1021 WTVNNRYSTWGG-----VSRLYVYPENYIDPTQIRIGTRMDELLNISQSKLSRDTV 1073  
Db 2116 FVQNSTFENANGGTLSPFAGNSLIIFAGNNHIAFTNHSGLTNLLSNQVSNINVTMLNASN- 2174  
Qy 1074 EDAPKTYLTRPETAADLVKVSAYHDVNSNTGLTWV----- 1110  
Db 2175 -----GLK-INATNNNSVSQ-NLFINASCVCQSQSDPTTASATNPTCTAQ 2217

Qy 1111 ---GQTRENLPEYYWRNVDIRMQ-----AGELAAANAKEMTKIDTAVNP-----YK 1154  
Db 2218 NNASSNASNAPIALNNDESLLVTANGFNFSNGIYANGVVDFSKIKGSANVKVLYLN 2277  
Qy 1155 DAIRPVIFRERLHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHDSGWSAPWSDIITQ 1214  
Db 2278 NA-----QOANNLTISNQAVLEKNA-----SFVTNLIQGAFFNNATQK 2318  
Qy 1215 VEAVTDKKPDERIALAASFGQGEDTLVFFV--KTGKSYSDFGSGNKNV----- 1262  
Db 2319 IEVL-----QNLVIASNA-----SLSTGIYGLVGGALNNLGAHFLENLSQTPVNPL 2366  
Qy 1263 -----AGMTIYGDSPFKOMENTALSY-----SOLKNTFDIHH 1295  
Db 2367 IQVGGLINLNTQTTPFMNVSVANGGTYYTLKLS---SRVIDYNINPNSLQSYLK-LYTLIN 2422  
Qy 1296 TQGNDLVRKASYRAQDFEVPASLNMGSAIGDSDLTVMENGINPOITSKYSSDNLAILH 1355  
Db 2423 INGNHIEKNG-----VLTYLQVRLLQDKGLL--SVALPNSNNAQNNILSLSVL 2472  
Qy 1356 NAAFTVRVYDGSNVRNKQISAMK--LTGVGDKSOY-----GNAPFIANTVKHYGY 1405  
Db 2473 HNQIKMSY---GNKVMDFPTPLQDIYVIGIOQSALNQIEAVGGNNAIKWLSTLAMWTK 2529  
Qy 1406 SDLGSPITVYKNTKNIYASVQGHLMADYTRLLILTPVENNYIARLFPPFPSPNTILNTV 1465  
Db 2530 NPLFAPIVLENHSLNEILGVTQDLQN---TASLISNPNFRNNATSLEW-----ASYTQ 2580  
Qy 1466 FTVGSNKTSPK-----KCSYAVDGNNSQGFIFSS-----YQSSG 1501  
Db 2581 QTSRLTKSLDPRAREGESNFSERLELKNKFPDNPSEVFVKYSQLSKHPNNLWQGVG 2640  
Qy 1502 WLDIDTGINNT-----DIKITVMAGSKTHFTTA--SDHIASLSPANSFDPMPYTFKP 1550  
Db 2641 GASPIGSGNGTLYGLNVGYDRLVKSVILGIVYVAGYSGFNGNIMHSLANNVDVGYARAF 2700  
Qy 1551 LEIDASSLAFTNNIAPLDIVFETKAKGRVIGKIKQTLV--KRVNYNP-----E 1598  
Db 2701 LKRNEFTLSAN-----ETYGNAHINSSNLSLVLNQRNYNTWTTSVNGNYG 2750  
Qy 1599 DILFIRET-----HSGAQYMQLV--YRILNTLLASQLVSRANTGIDTILTMETORUPE 1651  
Db 2751 DFMFKQKSVLKPQVGLSYHFIGLSGKMGKQNPAYQOFVMSHSPNSBSVLTLNMLGSSR 2810  
Qy 1652 PPLGEGFPANFVLPKYDPAEHGDERWFKIHGN-----VGGNTGROPYYSGLMLSDTSETSM 1707  
Db 2811 KYFGKNSY-----YFVTARLGRDLLIKAKGDNVVRFGENT--LLYRKGEIFNT----- 2857  
Qy 1708 TLFVYPAEGYYMH-----BQVRLGVGYQKI 1732  
Db 2858 --FASVITGEMHLWRLMYVNAVGLKMGLOVQDL 2890

## RESULT 7

US-11-052-554A-90  
; Sequence 90, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3.  
; SEQ ID NO 90  
; LENGTH: 3194

[illegible]

Db 2146 LS--NIVY-----EBSLNDNAIPKDLANNIPKDLGSKTSLSSLS 2182  
Qy 1871 P-----DAVQN-----DPMHYKIAFFMELLQOLIURG 1898  
Db 2183 PTEVNNLLGVSAFKVAIMEILNSKTVGVFGENGLLNALDPVKR-----EIDOMLLEQ 2236  
Qy 1899 DMAYRELTRDALNEAKMYVRLTELLGDEPDYGSQQAAPSLSGAASQTVQAAAYQODLT 1958  
Db 2237 IQAHS-----SGEKEPIVKTGLI--ENVENFINNWKYQKSLSSFANNFVFGGLNQALD 2287  
Qy 1959 MLRGGVSKNLRANSVLGLPEYNPALTDYQWTLRLRLFLNRLNSIDQPL--SLAI 2016  
Db 2288 KIGSSDAKIQS-----FLDK-----TFGDIL-----NOMINAPLNUKIS 2326  
Qy 2017 YAEPTDPKALITSMVQASQGSVALPGTSLYRFPVMLERTRNLVAQLTQFGTSLSSMAE 2076  
Db 2327 WLGPQDLSVNVNIALNSITNFSKELLGALS-----GMGQKVND-- 2365  
Qy 2077 HDDADELTTLLQOQMEIATOSIRIQORTVDEVDADI-----AVLAERSSAQN 2125  
Db 2366 -----LLGEGVNVKIMSNOVLGQWINKIADKGGVYHQGLGSILP--KSLQD 2412  
Qy 2126 PLEKYQQLYDEIDNHGEORAMSLIDAAQOSL--AGQVLSIAEGVADLVNPNVGLACGS 2183  
Db 2413 ELKK-----LGMG-----SLKPKGLHNLWQKGNFVAKOHVFNNSLSFNATGGE 2459  
Qy 2184 RWGAALRA-----SASVMSLSATASOYSADKISRSEAYRRRQEWIEIQRDN 2229  
Db 2460 LNFVAGKSIIFNGKNTINFTQYQGLSFVSKDFSNISLDTNATNGLTLNASKNDISVQK 2519  
Qy 2230 ADGEYKQMDAQLESUKIRREAAQOMVEYQETQQAHTQALQLLQKFTNKALYSMMRGKL 2289  
Db 2520 GQICNVNLD-----MTAKGKTTQTNSSSATAPTNETLEVSANFPA-----FLGTI 2566  
Qy 2290 SAIYYQFDLTQSCFLMAQEAALRELTONGTFFIRGGAWNGTTAGIMAGETILLINLAEME 2349  
Db 2567 KA-----NGLVDFSKVLQNTTIGTLDLGNATPK----- 2595  
Qy 2350 KVMLEDERALEVTRTVSLAQFYQALSSDNFNLTEKLFQFLREGKNGVAGSNE-----L 2404  
Db 2596 -----ANNLVNNAFNNNSYRANISGNFVAKGATFTNENGLNVGGNFSEGLPIF 2648  
Qy 2405 KLSNRQIBASVRLSLDKIFSDYPES--LGNTRQLKQVSVTL--PALVGPYEDIRAVLNY 2459  
Db 2649 NLNPNHTQIINVTGTSTIMSNQALINFNTQLKQGYATLINARMVYGV--DNQTL-- 2705  
Qy 2460 GGSIVMPGCCAIALSHGVNDSGQPLDFNDSRYLPFGISVNDSGSLTLPFD 2513  
Db 2706 GGSL-----SDYLKLYTLIDFNKRMQLNGDSLSYDNQFVSIKD--GGLVWSFKD 2753

## RESULT 8

US-11-051-453-42  
; Sequence 42, Application US/11051453  
; Publication No. US20050287150A1  
; GENERAL INFORMATION:  
; APPLICANT: AMBROSINO, DONNA  
; APPLICANT: BABCOCK, GREGORY J.  
; APPLICANT: BROERING, THERESA  
; APPLICANT: GRAZIANO, ROBERT  
; APPLICANT: HERNANDEZ, HECTOR JAVIER  
; APPLICANT: LOWY, ISRAEL  
; APPLICANT: MANDELL, ROBERT  
; APPLICANT: MOLRINE, DEBORAH  
; APPLICANT: THOMAS, JR., WILLIAM D.  
; APPLICANT: ZHANG, HUI-FEN  
; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: MJI-001  
; CURRENT APPLICATION NUMBER: US/11/051,453  
; CURRENT FILING DATE: 2005-02-04  
; PRIOR APPLICATION NUMBER: 60/542,357  
; PRIOR FILING DATE: 2004-02-06

; PRIOR APPLICATION NUMBER: 60/613,854  
; PRIOR FILING DATE: 2004-09-28  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn Ver. 3.3  
; SEQ ID NO 42  
; LENGTH: 2367  
; TYPE: PRT  
; ORGANISM: Clostridium difficile  
US-11-051-453-42

Query Match 1.5%; Score 201.5; DB 7; Length 2367;  
Best Local Similarity 19.1%; Pred. No. 0.0002;

Matches 382; Conservative 324; Mismatches 704; Indels 591; Gaps 110;

Qy 72 NPQLSGAIRLGIERSRSYDEMFGA-----RSSSEFV-KPGSVASM-FSPAGYLT----- 120  
Db 412 NDTLGPISQGNEDNTNNFSGESIGAIBENISFIAKIGSYLRVGFPEANTTTILSG 471  
Qy 121 -ELYREA-KDLHFSSAYHLNRRPDLADLTLSQSNMDTEISTLSN-ELLLEHITRKT 177  
Db 472 PTIYAGAYKDL-----LTFKEMSIDTSLSELRLNFEFPKVNISQAT 513  
Qy 178 GGDSDALMESLSTYRQALDTPYHQP-----ETIRQVIMTHDSTLSALSRLPEV 226  
Db 514 EOEKNSLWQFNREAKIQPEYKKNYPEGALGEDNDLDFSONTVTDKEYLLEKISSSTK- 572  
Qy 227 MQAEGASLLAILANISPELYNITLITEITKADALFAQNFSEMITPENFASQSWIAKY 286  
Db 573 --SSEGGVHVIV-----QLOGDKISYEAACNLFAKNPYDSILFORNIEDSEVAYY 622  
Qy 287 G---LELSEVQY-----LGMLOQNGYSD-STSAV-----VDNIST-----GL 319  
Db 623 NPTDSEIQEIDKYIPDRISDRPKIKLTFIGHGKAFFNTDIFAGLDVDSLSEIETAGL 682  
Qy 320 V-----VNNESKLEAYKITRVKTDYDKNINYFDLMYEGN--- 354  
Db 683 AKEDISPKSIEINLLGCMFYSVNVETYPCKULLRVK-----DKVSELMFSMSQDSIIV 738  
Qy 355 --NOFFIRANPKVREF-----GATLRKNAGSPGVGSLSGPLIANTNFKNYSLSIDSE 408  
Db 739 SANQYEVRLNSEGRELDDHSGEWINK-----ESIUKDISKEYISFNPKENKIIVKSKNL 795  
Qy 409 YKNGVKIYAVRYTSTATNOGGGIFTPESYPLTIFALKLNKAIL--CLTSGLSPLNQL 466  
Db 796 PELSTLLQEIIRNNSNS-----DIELEEKVMALECEINVISNIETQ 836  
Qy 467 TI-VRSNQAQIINDSV-LTKVFTLTFYSHRYAL-----SFDDAQLNGSVINOYADD 517  
Db 837 VVEERIEAKSLTSDSINYIKNEFKLIESIEALCDLKQONELEDSHFISFEDISETDEG 896  
Qy 518 DSVSHFNRLFTPLPKGKIFEADGNVTSIDPDEQSTFARGALMRGLGVNSGELYQLGKL 577  
Db 897 FSIRPINK-----ETGESIFV--ETEKTIFFEVA-----NHITEISKI 933  
Qy 578 AG-VLDAQNTTILSVFVSIYRLTLARVQL--TVNELCMYGLSPNG--KTASLSS 633  
Db 934 KGTIFDTVN-----GKLVKKVNLDTTHEVNTLNAAFPIQSLIENYSSKESLSLSV 984  
Qy 634 GELPLVILWYQ-----VTQWLTEAEITV-----BAIWLLCTPEPSGNISPEISNLN--N 682  
Db 985 AMKVQVYAQLFTGLNTITDAKVELVSTALDETIDLL--PTJSEGL-PIIATIIDGV 1041  
Qy 683 LRPSISDMAQSHNRELQAEILAPFIAATLHASPDMARYILLWTDNLRL-PGGLDI----- 737  
Db 1042 LGAAIKE--LSETSDPLLRQEIETAKIGMAVNLTTATTA-----IITSSIGIASGFSILLVP 1096  
Qy 738 -----AGPMTLVKESLNANETTLQV-FCHVMAQLSLSVQTL-----RLSEAEISLVLI 786  
Db 1097 LAGISAGIPSLVNNELVLRDKATKVVDYFKHVSIVETEGVFTLLDDKVMQOQDILVISEI 1156  
Qy 787 ---SGFAVLGA-----KNQPAQGH-----NIDTLFS-----LYRFHOWINGLGNPGSDTLQWL 831  
Db 1157 DFNNSNIVLGKCEIWRMEGGSGGHTVTDDIDHFFSAPSITYR-----BPHLSIYDVL 1207

```
Qy 832 RQQTTLTADRLASVMGLDISMVTQAMVSAGVNLQWODINTVLOWIDVASALHTMPSVIR 891
Db 1208 EVQKEELD-----LSKDLWLPNA-----PNRWFAWE-----TGW-----TPG-LR 1242
Qy 892 TLVNIRYVYALNKAESNLPWDEWQTLAENMEAGLSTQQAQ----- 932
Db 1243 SLEN-DGTKLLDRIRNDVGEFYYWRYFAFIADALITLKPRYEDTNRINLDSNTRSFIV 1301
Qy 933 --TLADYAEKLLSVLNCWFLANTQP-----EGVSLH-SRDDLYSYFLIDNOV-----S 978
Db 1302 PIITTEYIREKLSYFYSGGTYALPLSQYNGMINIELSESDW-IIDVDNVVADVITIES 1360
Qy 979 SAIKTRTLARAI-AGIOIYINR-ALNRIEPRARADVSTRQPFDTWNNRYSTWGWYS-- 1034
Db 1361 DKIKKGDILIEGILSTIEENKIILNSHEINFSEGVNGSGFVSLT-----FSILEGINAI 1416
Qy 1035 -----RLVYYPENYIDTPQIRIGQTRMDELLENISQSLSLSDTIVEDA 1076
Db 1417 IEVDLLSKSYKLLISGELKILMLNSHIQ--QKIDYIGFNSLQKNIPYSPVDSEKENG 1474
Qy 1077 PKTYLTR-----FETVADLKVVS-AYHDNVNSNTGLTWFGQTRNRLPEYKWRNVDSIRMQ 1131
Db 1475 FINGSTKEGLFVSELPLDVVLISKVYMDSDKPSFG-----YYSNNL----- 1514
Qy 1132 AGELAAANAKWETKIDTAVNPYKDAIRPVIFRERLHLI--WVEKEEVAKNGTDPVETYDR 1189
Db 1515 -----KDV--KVITKQNVNLTGYLLKDDIKISLS--LTIQDE 1548
Qy 1190 FTLKLAFLRHDSWSPWSYDITQVEAVTDKPDTERLALAAAGFOCEDTLLVYVYKGT 1249
Db 1549 KTIKLSVHLDESVA-----EILKFWNRKGSNTSDSLMSFLSMMNIKSIFVNFLO 1600
Qy 1250 KSYSDFGSNKNVAGMTYIGDGSFKKMENTALSYSQIKNTFDIHT--QGN--DLVRKA 1305
Db 1601 SNIKFILDANFIISGTTISGOFEEICDENNNIOQYFIKFNLTLETNYLYVNGRNQMVIEP 1660
Qy 1306 SY-----RPAQD--FVVPASLN--MGSALGDSDLVW-----ENGINIPOI-- 1341
Db 1661 NYDLDDSGDISSTVINFSQKLYGIDSCNVKVISPNYIDTINIPVYETNNYTPEVIV 1720
Qy 1342 -TSKYSNDNLAILTHNAFTVRY-----DGSGNVI-----RNKQISAMKLTGV----- 1383
Db 1721 LDANYINEKINVNIND--LSIRYVNSDNGDNDFILMSTSEENKVSQVKIRFVNVFKDKTLA 1778
Qy 1384 -----DGKSQYGNAPIANTVKHYG-----GYSDLGGPITVYKNTYNIASVQGHLMN 1431
Db 1779 NKLSFNFSKQDVPVSEIILSFTPSYEDGLIGY-DL-GLVSLYNE-KFYINNFGMMVSG 1835
Qy 1432 ADYTRRLILTPVENNYIARLPPEFPSP--NTILANTVFTVGSNKTSDFKCKSYAVD---G 1485
Db 1836 LIYIN-----DSLTY-----FKPPVNNLITGFTVVGDDK-----YYFNPIGG 1873
Qy 1486 NNSQGFQIFSS-----YOSSGWLDDTGTGINNTDKITVMAGSKTHTFTASDHIALPANSF 1541
Db 1874 AASIGETIIDDKNYFVNSGVL--QTGVFSTEDGKYFA-----PANTL 1915
Qy 1542 D-----AMPYTFKPLIEDASSLAFTNPIALDIVFETKAKDGRV-----LGKIKOTLS 1589
Db 1916 DENLEGAIDPTGK-LIIDENIYIFEDNYRG--AVEWKELDGMHYFSPETGK-----A 1966
Qy 1590 VKRYNYPEDILFLRETHSGAQYMQLVYRIRLNT--LLASQLVSRANTGTDITLTWTETQ 1647
Db 1967 FKGLNQIGDDKIYF-----NSDGVNQKGFVSINDNKHYPDDSGVMKVGYTEID----- 2014
Qy 1648 RLPEPPLGEGFPANFVLPKYDPAEHGDE-----RWFKIHIHNVGNTGQRPYYS 1696
Db 2015 -----GKHPIYF-----AENGEMQIGVENTEDGFKYFAHNNEDLNGEEGEBISYS 2058
Qy 1697 GMLSDETSETWTLFVPYABGYMHEGVLRGVYQKIYTDNTWESAFYFDE-TKQOQV-- 1753
Db 2059 GILNFNNKI-----YYFDSDFTAVVVGWKLDE-----DGSKYFYEDBTAEAVIGL 2102
```

```
Qy 1754 -LINDADH---DSCWMTQQGIV 1770
Db 2103 SLINDGOYYFNDNDGIMQVGFV 2123
RESULT 9
US-10-995-561-776
; Sequence 776, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 776
; LENGTH: 5935
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-995-561-776
```

```
Query Match 1.5%; Score 192; DB 6; Length 5935;
Best Local Similarity 17.7%; Pred. No. 0.0042;
Matches 485; Conservative 369; Mismatches 914; Indels 966; Gaps 123;
Qy 51 HETIEQKNNRL-----LEARIFTRANPQLS--GAIRLGIERSVRSYDEMFGARSSS 102
Db 410 HSYINVQNGQRLLLDKELMETLTSRDEYQTSPPKVVVEIGHQRQKTPE-----GLQESA 463
Qy 103 FVRPGSVASMPSPAGYLTLYREAKDLHFSASSAYHLNRRPDLADLTL---SQSNMDEI 159
Db 464 NVK---ISGTFS-SGWTVRL---PEQFSSQ-----NKEYPDREDCTTEKGGKTTVETED 511
Qy 160 STL-TLSNELLLEHITKTKGSDSALME--SLSTYROIDTPYHQVETIRQVIMTHDST 216
Db 512 SSVENPQDQLFVEQKERNPNDALKVINKVLEFVORQLIGT-----QREDQT 558
Qy 217 LSALSRNPE-----VMQAEAGASLAILANISPELYNLT-----E 252
Db 559 AVSVRENASRGHLITIPPAEAGVPLV-----VDKVFSEVETPKKEHQLPNTSFTCQNE 613
Qy 253 EITEKNADALFAQNFSENITPENFASQSWIAKYGLEL-----SEVOKYL 297
Db 614 QAHTLEYITHDEYTGSGSHIKPQSKLQVQVKTLGIKLELKSFTDGNVHPLDKKEMLKKT 673
Qy 298 GMLQNGYSDSTSAYVDNISITGLVNVNNSKLEAYKITRVKTDYDKNINYPDLMYEGNNQF 357
Db 674 FLAKDHKEQEA--QNIAGSMMWSE-----KTDEED-----SGREI 709
Qy 358 FIRANPKVSRREFGATLRKNAGPSGIVGSLGPIANTFNKSNLYLSNISDSSEYKNGVKIYA 417
Db 710 FL-----SCSHPLELLEBEATLNLV----- 728
Qy 418 YRYTSSTSATNOGGIITFFESYPLTIFALKLNKAILRCLTSGLSPLNELOTIVRSDNAQGI 477
Db 729 -----SAQLLDGGIF---HEQTGOKLLINEAI-----SRGI 756
Qy 478 INDSVLTKVFTTLFYSHRYALSFPDDAQVNLG---SVINOVADDDSVSHFNRLFTNPLPK 533
Db 757 VPSHTAVKLMKEL-----NMQGFDSQTCBSLITTEEVINEGLMDEKLLH-NVLNADKAIS 811
Qy 534 GKIFEADGNTVSDPDDEQSTFARSALMRGLGVNSGELYQLGK-----LAGVL 581
Db 812 G-----VLDPERTQLCSVKDAVTVGL-----LDKETATRIILERQVVVTGGII 852
Qy 582 DAQNTITLSYFVFISSLYRLTLL-----ABVHQLTVNELCMLYGLSPNGKTT 628
Db 853 DLKRGKKVSVTLASTGLGLVDVADQPELINLEKASKGRDAEKTVRE-----RLISLQMETT 907
```



; APPLICANT: Sachdeva, et al.  
 ; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
 ; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
 ; FILE REFERENCE: 30853/40359A  
 ; CURRENT APPLICATION NUMBER: US/11/052.554A  
 ; CURRENT FILING DATE: 2005-02-07  
 ; PRIOR APPLICATION NUMBER: US 60/589,227  
 ; PRIOR FILING DATE: 2004-07-20  
 ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
 ; PRIOR FILING DATE: 2004-02-06  
 ; NUMBER OF SEQ ID NOS: 763  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 92  
 ; LENGTH: 2399  
 ; TYPE: PRT  
 ; ORGANISM: Helicobacter pylori J99  
 ; US-11-052-554A-92

Query Match 1.5%; Score 191.5; DB 7; Length 2399;  
 Best Local Similarity 18.5%; Pred No. 0.00095;  
 Matches 423; Conservative 298; Mismatches 822; Indels 747; Gaps 108;

Qy	131	FSSAYHLDRPDLA-----DLTSLQSNMDTEISTLTSNELLEHITRTGG	179
Dd	385	YQSYLMDLINYKDKAISHVEVSNNYDVTDFDINGQDETTLQE-TFSNQSI---ITQPLG-	439
Qy	180	DSDALMESLTYRQADTPHPQYETIROVIMTHDSLTSLSRNPVWGQAG-	232
Dd	440	-DDLQQAQOQYQEDVANSQNALKVASDNTIANNDTSYTQSSNPILKDAQGLNTNQO	498
Qy	233	-----ASLALFLANISPELYN-----ILTEITEKNADALPAQNFSENIPTENPASQ	279
Dd	499	IQDEKALEKDLQAIK-QLANSTTGFPNEQAFTQAKQEQDEQALQDENAFTEQEGLE	557
Qy	280	SWIAK-----YGLSEISEVQKYLGLM-QN-	302
Dd	558	QAIANAKHANPTNPPTSPPTPIKHTAPNTPPSQVPPPTPSQNLKPTNVWNGVYVLQNK	617
Qy	303	GYSSTSAVD-NIS--TGLVNNESKLEAKITRVKTDYDKINYPDLMEQGNOFFI	359
Dd	618	TSYKNGIYYIDPNLSGSGQSGNTLSTYTANLLGR-----SFGVNANNGTLII-GNNTESV	672
Qy	360	RANFKVSRFCATLRKNAGPSGIVGSLSGPLIANTNFKSNYSISDSEYKGVKIYAYR	419
Dd	673	NDN-----GLIWGHGGFYITGTTSAANI-----YLTN-----NFKTG-----	706
Qy	420	YTSSTATNOGGGIFTFESYPLTIFALKLNKAIKRLCLTSLGSPNELOTIIVRSNDAQIIN	479
Dd	707	-EGVNSDGGGANITP-----KASDNITMDGLNYYNAETV-----	740
Qy	480	DSULTKVFTYTLFYSHRYALSFDDAQVNLGNSVINOYADDDSVSHF-----NRLFTPPPLKG	534
Dd	741	----TKMIQTGASQHSYT-TFDTATN--NISVTDSDFDSDMTWGFSPSAKNSISFNASFSG	793
Qy	535	KIFEADGNTVSDIPDERQS--TFEASALMRGLGVNSGELYQLGKLAGVLDQNTI---TSL	590
Dd	794	---FTNPGGSSTISVANSLSFTDSRL-----NGGAIYN-----LQANSLIFNNTQA	838
Qy	591	VFVI-----SSLYRLTLRLARVHQLTVNELCMLYGLSPFGKTKTASISGSGELPRLVIWLY	644
Dd	839	VFNVLVSRGTSNFNATTLQNGTSFTLSQSL-----NFGNDTTLQNNAN-----	884
Qy	645	QVTQWLTEAETITTAIWLLCTFPESGNSIPESINLNNLRPSISEDMA--QSHNRELQAE	702
Dd	885	-----ITLGNKSQAQAFKXNSLTLDNNSLSDQNSVLNANGTSAPFNQAASLNIYNGSQAAF	939
Qy	703	ILAPFIAATLHLAGSPDMARYILLTMDLRPGGLDIAGFTWLVLKES--LNANETTLQVQFC	761
Dd	940	SSLFNGGTLISNA-----NSKLNASSAFSNNNTIINDSDVLNANT-----	982
Qy	762	HVMAQLSLSVQTLRLSEABLVLVIGFAVLGAKNQAPAGQHNDITLPSLYRPHOWINGLG	821
Dd	983	---SSLNANINFGASQADF-----GGNTTIOT--ASFNF-----	1012

Qy	822	NPGSDTLMDLRQOTLTADRLASVMGLDISMVTQAMVSAGYNQLQCWODINTVLQWIDVAS	881
Dd	1013	-----DSASSLNFNNLTANGALNFNGYAPSLTKALMNVSGQFVLGNGDINI-----SDINI	1064
Qy	882	ALHTMPSVIRTLVNI-RYVTALANKA-----ESNLPFW-----DE	914
Dd	1065	FDNITKSVTYNILNAQKGIITGISCANGYEKILFYGMKIQATYSDNNNIQTWSFINPLNS	1124
Qy	915	WQTLAENNEAGLSTQQAQTLADYTAERLSSV-LCNWFLANIQPEGVLSHRDDL-----	967
Dd	1125	SQIIQESIKNG-----DLTIEVLNPNASNTIFNIAPE---LYNQDSKQNPQTG	1171
Qy	968	YSYFLIDNOVSSAIIKTRL-----AEAATAGIQLYINRALNRTEPNARADVSTRQF	1017
Dd	1172	YSYSDNOAQGTYYLTNSIKGLFTPKGSOPTPQPPGTYSPPNPQLNSI-----NIYKGF	1225
Qy	1018	FTDWTNNRYSTWGVSRVLYPENYIDPTQRIQOTRMDDELLENISQSLSRDTVEDAF	1077
Dd	1226	SSE-----NLKTLGILSONSATLKEMIESNQ-LDNITNINEVLQLLOKIKITQ-----AQ	1275
Qy	1078	KTYLTRTETVADLK-----VVSAYHDNVSNTGLTWFGQOTRENLPYWRNV	1125
Dd	1276	KQAL--LETINHLTDNINOTFNNGNLVIGATQDNVTNMTSSIWFGG-----	1319
Qy	1126	DISRMQAGELAAWKEWTKIDTAV-----NPYKDAIRPVIERRERLHLIWVEKEEVAKN-	1179
Dd	1320	-----NGYSFPCALDSATCSSFRNTY---LGQLGSTSPYLGVINADFKAKSI	1364
Qy	1180	---GTDPVETYDRFTLKLAFLRHDSWSAPWSYDITTVQBEAVTDKDPDTERLAALASGFG	1236
Dd	1365	YITGT-----IGSSNAFESGGSADVTFQ-----SANNVLNKANIE	1400
Qy	1237	GEDTLVVFVYTKSVSDFGSKNKNVAGM-----TIYDGSFKK-MENTALSRVSQIKNTF	1291
Dd	1401	AQATDNIF-----NLLQGEIDKIFNOGNLANVLSQWAMEKIKQACGLG	1444
Qy	1292	DIHTQCNDLVRKASYFAQDFEVPASL---NMGSAIGDDSLTWMEN-----	1335
Dd	1445	NFIENALSPLSK-----ELPASQDETGLQLGQNNLDDLNNSGVMNEIQNIS	1494
Qy	1336	-----GNI--POITSKY-SSDNLAITLHNA--FTVRYDGGNV-----TRNK	1373
Dd	1495	QKLSIFGNFVTPSIENYLAQSKSLMDKGLNLFIGGYIDASELSILGVILKDTNP	1554
Qy	1374	QISAMKLTGVDGSKQYGNAFIIANTVK---HYGYSDL-----GGPITVYNTKRYI-	1422
Dd	1555	PTSLQKDIGVANDLL-NEFLGQDVVKLESQGLVSNIIINNVSQGLSGVYNGGLSVL	1613
Qy	1423	-ASVQGHLMNADYTRRLILTP-----VENNYARLFFPPFPSP-----NTIL	1462
Dd	1614	PPSLQNALKNDL--GTLLSPRGLHDFWQKGYFNFLSNGVTVVNNSSFSNATGSLNFVA	1671
Qy	1463	NTVFTVGSNKTSPFKCSYAVDGNNSQGFQIPSS-----YQSSGMLDIDTGINNT	1512
Dd	1672	NKSIIFNGDNTIDFSK-----YQGALIFASNGVSNINITTLNATNGLSLNAGLNV	1722
Qy	1513	DIKITWAGSKTFTFASDHASLPANSPAMPYTFKPLDIDASSLAFTNNAIPLDIVFE	1572
Dd	1723	SVQKEICINLANCPPTTKN---SSPANS--SVTPPTNESLSVHANNFTFLGTIISNGAIDL	1777
Qy	1573	TKAKDGRVLKIKITLSVKRVNYPEDILFLRETHSGAQYMQLVGYRILNTLIASQL--	1630
Dd	1778	SQVTNNSVIGTL-----NINENA-----TLQANNLTI	1804
Qy	1631	-----VSRANTGIDTILTMETQ-RLPEPPLGEGGFANFVLPKYDPAEHGDERFKIH--	1691
Dd	1805	TNAFNNASSTANIDGNFTLNQOATLSTNAGLVMGNF-----NSYGDVFNLSHVS	1857
Qy	1682	---IGNVGN---TGRQPYYSGLSDTSETMTLFPVYAEYGMHEGVLGVGY-QKIT	1733
Dd	1858	SHAIINTQGTATTAMNNPLIQFNASSKEVGTVTLIDSAKAIYY-----GYNNOIT	1908





```
Db 1266 -----DKTRAASIVDVL SAGFNLQNGEAVDFVSTYDVTNFANGNTTAKVTY-- 1313
Qy 1199 HDGNSAPWSYDIT--TQVEAVDTKPDTERLALAAAGFOGEDTLTVFVYKTKSY-- 1252
Db 1314 DDTSKTSKVYVDVNVDDTTIE-VKDKKGLGVKTTTLTSTG-TGANKFALSNOATGDALVKA 1371
Qy 1253 SDF-----GGSNKVNAGWTIYDGSFKKMENTALSRYLSOLKN--TFD--- 1292
Db 1372 SDIVAHNLTLSGDIQTAKGASOANSAGYVDADGN-KVIYDSTONKYFOAKNDGTVDKTK 1430
Qy 1293 -----IHTQGNLVRKASYR-----FAQDFEVPASLN-- 1320
Db 1431 EVAKDKLVAQAQTPDGTGLAQNVRKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNKT 1490
Qy 1321 MGSAL-----GDSLTFWNGNIPQI-----TSKYSSDMALITLHNAA 1358
Db 1491 KNAAVTVGDLNVAQAQTPLTFAAGDTGTTAKKLGETLTKGGQTDNKLTDNNIGVVAGTDG 1550
Qy 1359 FTVR-----YDGSNVIRNKQISAMKLTGVGDKSOYGNAPIIANTVKHYGYS DL 1408
Db 1551 FTVKAKDLTWLNSVAGGTIKDEKGIS---FVDANGQAKANTPVLGANGL-----DL 1600
Qy 1409 GGPITVYNKTKNYIASVQGHLMNADYTRRLILTPVEN--NYIARLFPPFPSPN----- 1459
Db 1601 GGVV-INSVVGKTKDQDAANVQQLNEVNLGLGNDADGNQVNIADIKKDPNSGSSNR 1659
Qy 1460 TILNTVTVGNSKTSDFPKCSYAVDGNNSQFQJFSSYQSGGWLDDTGIN-----NTDIK 1515
Db 1660 TVIRAGTVLGGKGNNDTEKLA-----TGGVQV-----GVDKGNANGDLSNVWVK 1704
Qy 1516 ITVMAGSK---THFTA---SDHIASLPANSFDMPTFKPLEYDASSLAP--TNNTIAPL 1567
Db 1705 -TQDGGSKALLATYNAGQTNVYNNPFAEID-----RINEQIRFFHVNDGNQE 1754
Qy 1568 DIVPETRAKQGRVLKIKQTLKSVKRVNVPNPDIILFLRETHSGAQMOLG 1616
Db 1755 PVVQGRNGIDSSASGKHSVAIGFOAKADGEAAVAIGRQTQAGNSIAIG 1803

RESULT 12
US-11-102-476-46
; Sequence 46, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U. M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102,476
; CURRENT FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1889
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-46

Query Match 1.4%; Score 178.5; DB 7; Length 1889;
Best Local Similarity 18.8%; Pred. No. 0.0046;
Matches 377; Conservative 255; Mismatches 671; Indels 701; Gaps 95;

Qy 7 LNKISPTRDQGTWTLADQLVLSSEL-----RKIPDQLSWGCEARHLYHETIRQKKNR 61
Db 226 VLSQIQTNNLNPGLGKKQKLTATGIYSDNSNRDSSSVNWS---NSTIANQNG 281
Qy 62 LLEARIETRANPOLSGAIRLGIERSVRSYDEMFGARSSSFVKPGSVASM-FSPAGYLT 120
```

```
Db 282 ILETAD-----TGIV-----TVSASTENIIGS-VKLIIVTPAALVSVSPSTNSTV 325
Qy 121 ELYREAKDL--HFSSSAYHLNRRPDLAD-LTILSQSNMDT-----EISTLTLSN 166
Db 326 -----AKGLOENFKATGIFTDNSNSDITDQVTDWSSNTDILSISNASDSHGSLASTLNOGN 380
Qy 167 ELLLEHILTRKGG-----DSDALMESL-----STYROAIDTFPHQPYETI-- 206
Db 381 ----VKVTASIGGIQGGTDFKVTOEVLTSIEVSPTRISIAKGLTKQKTAIGIFTDNSKDD 436
Qy 207 --RQVIMTHDSTLSALSARNPE-----VMQOAGASLL-----AILANI- 242
Db 437 ITDQVTWNSSSAIVSVSNLDDNKGKGAHAVGDDTTITATLGKAVAGKTWLTVPVAVLSIQ 496
Qy 243 ----SPELYMLTBEETEKADALFAONFSENITPEN--FASQSWIAKYVLELSEVQKY 296
Db 497 INPVNPSLAKGLTKQFT---ATGIYSDNSNKDITSAVTFWSSDSSIA-----TISNAQKN 548
Qy 297 LGMLQNGYSSTSAVDNISITGLVNNESKL-----EAYKI 332
Db 549 QG---NAYGAATGTTDIKATFGKVSSPVSTLSTAAKLVEIQITPAAASKAKGLTERFKA 605
Qy 333 TRVKTDYDKNINYFDLMYEGNNQOFFIRANFKVSRFEGATLRKNAGPSGVISLGSGLIJA 392
Db 606 TGIFTDNSNDI-----TNQ-----VTWNSSNTDIAEIKNTSGSGKITNTLT---P 648
Qy 393 NTNPKSNVNLNISDSSEYKNGVKIYAYRYTSTSATNOGGGIFTFESPLTIFALKLNKA- 451
Db 649 GSSEISAALGSIKSSK-----VILKVTTPAQ 673
Qy 452 -IRLCSTGSLSPNELQTIIVSDNAQIINDSVLTKFVYTLFYSHRYALSFDDAQLN--G 508
Db 674 LISIATVP-INPSVAKGLIRQFRATGYTDHISVQDVTALATWS---SSNPGKANVNVVG 729
Qy 509 SV-----INOXADDSDSVSHFNLFTNPLKGIKIFEADGNTVSIIDPDEEOSTFARSALM 561
Db 730 SVTTVATGNTNIKATIDISIGSSSVLNVTPALL-----TSIBITPTINSITHGLTKQF 781
Qy 562 RGLGVNSGELYQLGKLAGVLDQAQNTITLSVVFVSSLYRLTLARVHQLTVNELCMLVGLS 621
Db 782 KATGIFSDK-----STQNLTLQTLVTVISSDPSKIE-----IENTS 815
Qy 622 PFNGKTTAS-LSSGELPRLVILWYQVTOQLTEABITTEAIWLLCTPFSGNISPEISNLL 680
Db 816 GKGIATASKLGSNNIKAVYKFSQSPPIPIVTDLKLKSI-----TISSSSSIA 865
Qy 681 NNLRPSISEDMAOSHNRLOAEILAPFAATLHLASPDMARYILLWTDNLRPGGLDIAGF 740
Db 866 KGL-----TQQFKAIGTFIDG- 881
Qy 741 MTLVLKESLANETTQLVQFCHVMAQLSLSVQTLRLSEABLVLVSGFAVLGAKNQAPAG 800
Db 882 -----SEQETINLV-----TWYSSKSDVAPINNAAKGLATALSIG 918
Qy 801 QHNIDTLFSLYRPHOWINGLNGPNSDGLDMLRQOQTLADRLASVMGLDISMVTQAMVSAG 860
Db 919 SSNLSAIYNSISNK-IN-FNVSAAITLDSIKINPVN-NNIAK--GL-----TQOYTALG 967
Qy 861 VNQLQCMQDINTVLQWIDVASALHTMPSVIRTLVNIRYVTALNKAESNLPFSWDEWQTLAE 920
Db 968 VYSDTIQDISDSVTM---SSSNSSSISISNSTETKGKATALQIGKSKITA--TYSNISE 1022
Qy 921 NMEAGLSTQQAQTLADYTAERL-----SSVLNWFANIQPEGVSLHSRDDLYSYFL--- 972
Db 1023 NIDI---TVSAAATLSSISISIPINTINATVSKQPFAMGTYSYG---TKADLTSSVTWSS 1075
Qy 973 ---IDNOVSSAIKTRTALAEAI-----GIOLYINRLNRPEPNARADVS- 1013
Db 1076 SNKSQSKVSNASKTKGLVTGIASGNSIITATYGSVSGNTILTVNKT-DTAPTQVSUSSL 1134
Qy 1014 ---TRQFFTQVTVNNRYSTWGGVSRL-VYYPENYI-----DPTQIRIGOTRMMDELLENISQ 1065
```



Db 1135 SPPTIQVYVSINNKEAL--DLSNKYIINGSNFIGHCSNDTDFNSNQADFLSSIKG 1192  
 QY 1066 SK-----LSRDTVDAPKTYLTRPETVADLKVVSAYHDVNSNTGLTFWVGQTRNLPE 1119  
 Db 1193 SKNFTTILSHQILINKSYTLVNVKQGHDLSSI-----PNSLSCPNNSDFIG----- 1240  
 QY 1120 YYWRNVDISRMOAGELAAANAKWKTKIDAV-NPYKDAIRPVI FRERLHLIWEKEEVAK 1178  
 Db 1241 -----KEQLKLTSAVCNSLNQVI--VSFSKPLY-----SGKEATK 1273  
 QY 1179 N--GTDVETVDRFTLKLAFLEHDSWSPWSYDITTOVEAVTDK-----KEDTERLALA 1231  
 Db 1274 SVECSNPSQCESRY--KEAGVSSLSG-----ITSVRILDKGVCGGAPADSSKICLT 1322  
 QY 1232 ASGFGEDTLLVYVYKTSYDFGGSKNVAGMTIYDGSFKKMENTALSRYSLKNTF 1291  
 Db 1323 HSLLOS-----GGQYIIAANDLNGDGNK-----SWGAIKDSF 1357  
 QY 1292 DIIHQ-----GNDLVRKASY----- 1307  
 Db 1358 DOENLQSPKDKRINFICGNSPLNFMGPIVSDPFGDSDGFLVDYNNQIYLGPNVKN 1417  
 QY 1308 ---RPAQDFEVPASL--NWGSAIGDSDSLTWMEGNIPQITSKYSDNLAILHNAFTV- 1361  
 Db 1418 QAARFNYDGTTPESIFFSTQDINATNRASRDGGIP-VPNYVTIHTGCTILNSADITG 1476  
 QY 1362 ---RYDGS-----NVIRNKQISAMKLTGVD 1384  
 Db 1477 GPDNEDGRGVATGSLDKSHIFLAGSKPSFNYLYSSDITDNLNFKYISMGKITGL- 1535  
 QY 1385 GKSYGNAFIANTVKHYGYSGLGPIVYVYKTKNYIASVQGHLM--NADYTRRLILTP 1442  
 Db 1536 --ATAGTSSIAVLDDRIHVGA-----KKNQNLNAPDPFGKITFNTSEHNRCAIVNN 1584  
 QY 1443 VE-----NNYARLPFEPF-----SPNTILNTVFTVGSNKTSDFKKCS 1480  
 Db 1585 CEASDGYRGNRFP--RIDRMPYFGGSDVDVNYRYSKSDNSINWGYVIGDLSLFFVKSL 1642  
 QY 1481 YAVDG--NNSQGFIFSSYOS-----SGWLDITGINTDIKITVMAGSKTHT 1526  
 Db 1643 YAANGFPNSLNGSIHISTSANPSPCEGINRCSWKDTAPSN-----PKWH- 1690  
 QY 1527 FTASDHASLPANSDAMPYTFKPLEIDASSIAFTNNIAPLDIVPETKAK-DGRVLGKIK 1585  
 Db 1691 -----NSPHNNFSL-----ELTKYRNLIIPADKAPSPQAFENGRL----- 1725  
 QY 1586 QTLVSRVNYNPEDILFRETSHGAYQYQMLGVYRIRLNTLASQLVSRAVTGIDITLME 1645  
 Db 1726 -----YVTRTICVTKEHSGLRQ-----SLQTVKGTGD-----SY 1756  
 QY 1646 TORLPEPPLGEGFFANFVLPKYDPAEHGD-----ERWFKIHIGNVGN-TGRQPYSGM 1698  
 Db 1757 TNRRPQ-----LWKCDDPILTGDTTCEADW-----SLVGNGTGFTNF----- 1795  
 QY 1699 LSDTSETSWTLFPVPAEYGMHEG 1722  
 Db 1796 -GDNNSHNSMTMV--ASGSYLVI 1816

RESULT 13  
 US-11-075-185-28  
 ; Sequence 28, Application US/11075185  
 ; Publication No. US200502664341  
 ; GENERAL INFORMATION:  
 ; APPLICANT: REEVES, CHRISTOPHER D  
 ; APPLICANT: JULIEN, BRYAN  
 ; APPLICANT: REID, RALPH  
 ; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS  
 ; FILE REFERENCE: 010099.03  
 ; CURRENT APPLICATION NUMBER: US/11/075,185  
 ; CURRENT FILING DATE: 2005-03-07  
 ; PRIOR APPLICATION NUMBER: US 60/551,103  
 ; PRIOR FILING DATE: 2004-03-08

; PRIOR APPLICATION NUMBER: US 60/568,290  
 ; PRIOR FILING DATE: 2004-05-04  
 ; NUMBER OF SEQ ID NOS: 61  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 28  
 ; TYPE: PRT  
 ; ORGANISM: Sorangium cellulosum  
 ; US-11-075-185-28

Query Match 1.3%; Score 170.5; DB 7; Length 2087;  
 Best Local Similarity 18.2%; Pred No. 0.019;  
 Matches 406; Conservative 304; Mismatches 860; Indels 661; Gaps 103;

QY 532 LKGIKFEADGNTVSDPDEQSTFARSALMRLGLVNS--GELYQIGKLAGVLDAQNTITL 589  
 Db 128 LDGIADTFNDVEMN--EKFAPELERLSRAVGKEGIGQVSMGEVSGAW-ADEVASV 183  
 QY 590 SVFVISSLYRLTLARVHQLTV-----NELCMLVCLSPFNKTTASLSSGELPRLVWLY 644  
 Db 184 NALIGDLVQPTREMARVIGAVAKGDLSTQMALEVGRLEGE--FLQTAQTVNTV--Y 238  
 QY 645 QVTQWLTEA-----BITTE-----AIWLLCTP--EFSGNISPEISLNNL 683  
 Db 239 QLGSFASVTRVAREVTEGKLGGOAEVKGAVGTWKDLTDSVNSMAGNLTAQVRNIAVT 298  
 QY 684 RPSISEDMAQSHNRELQAEIL-----A 705  
 Db 299 TAVANGDLTKITVDVVRGEILELKDFTMTVDQLRSPASEVTRVAREVTEGKLGQASV 358  
 QY 706 PFIAATLHLASPDMARYILLMTDLNLRPGGLDIAGFMTLVLKESLNANETTQLV-QFCHVM 764  
 Db 359 PGVAGTW---KDLTDSVNSMASNLTAQVRNIAAVTTAVANGDLTKITVDVKGILELK 414  
 QY 765 AQLSLSVOTLRLSABLSVLVSGFAVLGAKNOPAQHNIIDTLPFLSRFHQWINGLNP 824  
 Db 415 DTFMTVDQLRSPASEVTRVA---REVGTGKLGQAE-----VKVAGTW 457  
 QY 825 SDTLDMLRQTLTADRLASVNLGLDISMTQAMVSAGVNLQOCQWQDINTVLOWIDVASALH 884  
 Db 458 KDLTDSV--NSMASNLTAQVR--NIAAVTTAVARGDLTQ-KITVDVVRGEI--LELKDTFN 510  
 QY 885 TWPVSIRTLVN-IRYVTLAKAESNLPWDE-----WOTLAENWEAGLSLQOQOTLAD 936  
 Db 511 TWDQLRSFASVTRVAREVTEGKLGGOAEVKGAVGTWKDLTDSVNSMASNLTVQ-LRD 569  
 QY 937 YTABRLSVLCNWFEL--NIQPEGVSLHSDLDLYSYFLIDNOVSALIKTRLA----- 987  
 Db 570 --VSKVATAIANGDLTKITVDVVRGEILELQIKDVINT--TVDLSSFAAEVTRVARDVGE 625  
 QY 988 -----EAIAGIQLYINRALNRIEPNADVSTROFTDWTVNNRYSTWGGVSRVLY 1038  
 Db 626 GKLGQOAEVKGAVGTWKDLTDSVNSMASNLTAQ-----VRNIAAVTTAVAR--- 671  
 QY 1039 YPENVIDPQIRIGOTRMDELLENISQSLSRDVEDAFKTYLTFTFETVADLKVVSAYHD 1098  
 Db 672 ---GDLTKI-TVDVRGEILE---LKNFTNTMVDQLRSFAAQVTRVA----- 711  
 QY 1099 NVNSNTGLTWFGVGTRENLPEYTWYNDI SRMQAGEL---AANAWKSWTKIDTAVNPK 1154  
 Db 712 -----REVTEGKLGQOAEVGTGAGTWKDLT----- 737  
 QY 1155 DAIRPVI FRERLHLIWEKEEVAKNGTDPVETYDRTLLKLAFLRHGDSWSPWSYDI--- 1211  
 Db 738 DSV-----NSMASNLTAQVRNIAVDTTAVA-----NGDLSKKITVDVRGE 777  
 QY 1212 -----TTQVEAVTDKPDTERLA--LASGFGEDTLLVYVYKTKSYSDFGSN-- 1259  
 Db 778 ILELKDFTMTVDQLRSPASEVTRVAREVTEGKLGGOASVPGVAGTWKDLTDSVNSMAS 837  
 QY 1260 -----KNVAGWT---IYDGSFKKMENTALSRYSLKNTTPTDI IHTQGNLVRKASYPFA 1310  
 Db 838 NLTAQVRNIAVDTTAVARGDLS--KKITVDVKGILELKNTP-----NTWVDQLSSFAA 889

Qy 1311 QDFVPASLWMSAIGDSSLTWEMNGNIPOITSKVSS--DNLAITLHN-AAFT----- 1360  
Db 890 EVTRVAREVGTGKLGQAQAEVTVAGTGWKDLTDSVNSMASNLTAQVRNIAAIVTAVANGD 949  
Qy 1361 ---VRDGSNGVIRNKOISAMKLTGVGDKSQYGNAPI--TANTVKHGGYSDLGGPITV 1414  
Db 950 LSKKITVDVRGEIIELEK--NTINNTWDDQ-----NAPASEVTRVAREVGTGKLGQAQSV 1003  
Qy 1415 -----YNTKTKNYIASVQGHLMNADYTRRLILTPVNNYIARLFEPPF-----SPN 1459  
Db 1004 PGVAGTGWKDLTDNVFNAGNLTNQVRGIKAVVTVAVANGDLKRLAFDAKGAIAALADTIN 1063  
Qy 1460 TILNTVTVGSKTSDPKKCSYAVDGNNSQGFIFSSVQSSGWLDDITGINN----- 1511  
Db 1064 GVIEITLAFADQVTVVARE--VGVEG--KLGQAQSVGAAGTGWKDLTDNVNQLAANITQ 1119  
Qy 1512 ---TDIKITVMAGSKTHTF--TASDHASLPANSFDAMPYTFKPLEIDASSLAFTNNIA 1565  
Db 1120 VRAIAEVATAVKGDLTFTIRVEAQGEVASL--KDTINEMIRNLKDTYLNKSE----- 1170  
Qy 1566 PLDIVFTKADGKVLGKIKOTLSVKRVNPNEDILFLRETHSGAQYQMLGVYRIRLNTL 1625  
Db 1171 -QDWLKTNLAKFSRLQOKDLLTQVR----- 1197  
Qy 1626 LASOLVSRANTGIDTILTMETORLPEPPLGEGFPANFVLPKYDPAEHGDERWPKIHIGNV 1685  
Db 1198 ILSELAPVVGAAQGVFFTMVAK--EPTIL--KLLASYA---YKVRKGVNDNH--FKLGEGLV 1250  
Qy 1686 GGNTGRQPYSGMLSDTSETMTLFPVPAEGYVMHGVRLGVGYOKITYDNTWESAFYF 1745  
Db 1351 G-----QCAL----- 1255  
Qy 1746 DETKQFVLINDADHDSGMTQOGIVKNIKKYGFANVSIAATGYSPMDFNSALYYWEL 1805  
Db 1256 --EKEKILLVN-APPD-----YIRITSGLGEAPPVNIIVPLFEGQV 1295  
Qy 1806 FYTTPMCFQRLLEKQ--PDEATQWVYVNPAGIYVNGEIAPIWNCRPLETTSWNA 1863  
Db 1296 KAVIELASFERFSPHQAFLDQLESIGIVLNTTEANNRTE-----DLLKQOSLSA 1346  
Qy 1864 NPLDAIDPDVAQNDPMHYKIATFMRLLDQLILGDMAYREL--TRDALNEAKMYVRTL 1921  
Db 1347 RELSQOEELQOTNAELGEK-----ARLLAQONVEVERKNGVEQAQALEEK----- 1394  
Qy 1922 ELLGDEPEDYGSQOAWAPSLGASQTVQAAIQDQLTWLGRGGVSKNLRT-ANSLVLGLFL 1980  
Db 1395 -----ARQLAITSKYKSEFL---ANMSHELRTPLNSLL----- 1424  
Qy 1981 PEYNPALTDYQTLURLRLFNLRHNLSDGQPLSLA--IYAEPTDPKALLTSMVQAS--QG 2036  
Db 1425 -----ILSDQ-----LSKNDRNRTGRQVEFAKTHSGSNDLLALINDILDLSKIES 1471  
Qy 2037 GSAVLP--GTLSLYRFPVNLERTNLVAQLTOFGTSL-----LSMAEHDDADELTLLLLQ 2090  
Db 1472 GTVIDVGEISFSDLODYVERTFRHVAESKLEFELNFAQNLPOVIYTDKGVQOVL--- 1528  
Qy 2091 GMELATQSIRIQRTVDEVDAIDIAVLAEBSRRAQNLREKYQOQLYDEBINH--GEORAWSL 2149  
Db 1529 -KNLLSNSFKPTEGSRVALDVD--LVTSGWAPEN-----EGLSRAGAIAIAMSVR 1574  
Qy 2150 DAAG-----QSLAGQVLSIAEGVADLVPNVFGLACGSRWGAIRASASVM-----SLS 2199  
Db 1575 DTGIGIPHDKQOIIFEAQQADGST-----SRKYGTGLGLAISREIAMMLGGEIKLS 1627  
Qy 2200 ATASQYADKILSRSEAY-----RRRQEWEIFORDNADGEVKQMDAQLESIKTRREAAQOMVE 2256  
Db 1628 SKPGSSSFYLYPLTYTPARPRRKEQTVEPSPAPPVAVSGDVP-----PRSA----- 1676  
Qy 2257 YQETOQAHTQAQLELLQKFTNKALYSWMRG--KLUSAIYYQFPDITQSFCLMAQALRREL 2315  
Db 1677 -----BPPPHLLNQSDDSA-----GLKPSDSVVLIVENDASFAHFVMDVAH--- 1718

Qy 2316 TDNGVTFIRGAMNNTTAGLMAGB-----TLILNLAEMEKW--LERDERALEVTRTVS 2367  
Db 1719 -DHGFKAIL--AVRGGAALSIVERRRVAITLIDINLFDMDG--WRVLDRVRKDLB--TRHIP 1773  
Qy 2368 LAQFYQALSDNFNLTEKLTQFLRBGKGNVGAAG--NEKLKLSNRQIEASVRLSDLKIFS 2424  
Db 1774 V---QVITTD-----EERERALR-----MGAKGVLCPLKTRDALDETFRRLSQFMVSS 1819  
Qy 2425 -----DYPESIGNTRQLKQVSVTLPALVGPYEDIRAVLNYGGSIVMPRG 2469  
Db 1820 RRKIVLAPGDAERQELVELLGGD-----DVTIRSVASGEEALDALVTERPDVLLIR-- 1871  
Qy 2470 SAIALSHGVNDSGQFMDLDFDNDSRVLPPEGISVNDSSG-----LTLSPDADTDROKALIES 2524  
Db 1872 -----LDLPDVRCPDLIGLQAQSGSGTDLPLVLYVAPEEISPADEAQLSR 1915  
Qy 2525 LSDIIL--HIR 2533  
Db 1916 FSQLMWLKHVR 1926  
RESULT 14  
US-11-113-424-14  
; Sequence 14, Application US/11113424  
; Publication No. US20050260713A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli et al.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-225  
; CURRENT APPLICATION NUMBER: US/11/113,424  
; CURRENT FILING DATE: 2005-04-21  
; PRIOR APPLICATION NUMBER: 60/256,704  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/311,590  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/257,314  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/311,613  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/315,617  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/307,506  
; PRIOR FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: 60/322,358  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/294,075  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/288,153  
; PRIOR FILING DATE: 2001-05-02  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 2769  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-113-424-14

Query Match 1.38; Score 167.5; DB 7; Length 2769;  
Best Local Similarity 18.1%; Pred. No. 0.048;  
Matches 352; Conservative 244; Mismatches 598; Indels 755; Gaps 99;  
Qy 826 DTLDMLRQOITLADRLASVMGLDISMVTQAMVSAGVNLQOC-----WQDINTVLQWIDVA 880  
Db 1119 KDTDVYNNQK-----VFGL-----SEAFVSVGVEYESCPDILLKEKRTTVLQGYEID 1164  
Qy 881 SALHTMPSVIRTLVNIIRYVTTALNKAESNLPFSWDEWQTLAENMEAGL-----STQQ 930  
Db 1165 A-----SKJGWSLDDKHHALNIQSGILHKGNGENQFVSQ 1199  
Qy 931 AQTADYTA-----ERLSSVLCNW-----FLANIQPEG- 958  
Db 1200 PPVIGSIGMNGRRRSISPCSCNGLADGNKLLAPVALTCGSDGSLYGVDFNIRFPFSGN 1259

QY 959 -----VSL-----HSRDDLYSYFLIDNQVSSA-----IKTT-----RLAE 988  
Db 1260 VTNILELRKNDPGRSHSPAHKYLLATDPMSCGAVFLSDNSRVRFKIKSTVWVVKDVLKXNSE 1319  
QY 989 ATAGIQLYNRAL-----NRIEPNARADVSTROFFDWTNNRYSTWGVSRVLYVYPENYI 1044  
Db 1320 VVAGTG-----DQCLPDDTRCGDGGKA-----TEATLTNPRGTVTDKFGLIY-----FV 1364  
QY 1045 DPT--QRICQTRMDDELL-----ENISQSKLSRDTVEDAFKTYLTRPETVADLKVVSAYHDN 1099  
Db 1365 DGTMRIRIDONGIISTLLGSNDLTASRPLSCDSV----- 1398  
QY 1100 VNSNTGLTFVCGQTRNLEPEYFWRNVDSRMOAGELAAWKEWTKIDTAVNPYKDAIRP 1159  
Db 1399 -----MDISQVRL-----EW-PTDLAINPMDNSL-- 1421  
QY 1160 VIFRERLHIWEKEVAKNGTDP-----VETVDRETL-KLAF-----LHDSGW 1203  
Db 1422 YVLDNNVVLQISENHQVRIVAGRPMHCQVPGIDHFLSKVAIHATLESATALAVSHNG-- 1479  
QY 1204 SAPWSYDITTVQEAVTDKKP-----DTERLALAASG-----FGQEDTL 1241  
Db 1480 -----VLVIAETDEKINRIQVTTSGEISLVAGAPSGCDCKNDANDCDFSGDDGY 1530  
QY 1242 L-----VYVYKTKSY-SDPGG-----SNKNVAGWTIYDGSFKKME 1277  
Db 1531 AKDAKLNTFSLAVCAGDELYVADLGNIRIRIRKNKPFPLNTQNNYELSSPIDQELYLF 1590  
QY 1278 NTALSRYSQKNTFPIIHTQGNDLVRKASYRPAQDFEVPASLNMSAIGDSDLTWMNGN 1337  
Db 1591 TTGKHLTYQSLPTGDIY-----NFTYGD-----GDITLTIDNNGN 1627  
QY 1338 IPQITSKYSSDNLAITLHNAFTVRYDGSN-----VIRNKQT-----SAMLKTGYDG 1385  
Db 1628 M-----VNVREDSTGMLVLVVPDGVYVWMTGNTSALKSVTTQG 1667  
QY 1386 -----KSQVGNAPIIANTVKHYG-----GYSDLG-----GPITYNKTKNYIASVQ 1426  
Db 1668 HELAMMTYHSGILLATKSNENGWTTVEYDSFGRLTNVTFPTQGVSSFRSDTSDSVHVQ 1727  
QY 1427 GHLMNAD-----YTRBLILTPVNNY-----ARLF-----EPFSPN 1459  
Db 1728 VETSKDDVTITNLSASGAFYT-----LLOQVNRNSYIIGADGSLRLLLLANGMEVALQTEPH 1785  
QY 1460 TILNTVF--TVGSKNTSDFKKCSYAVDGNNSQGFQIFSSYQSGWLDIDTGINNTDIK--- 1515  
Db 1786 LLAGTVPNTVGRNVT-----LPIDNGLNLVEMRQK 1817  
QY 1516 -----ITVMAGSKTHFTASDHIAASLPANSPDM-----PYTFKPLETDASSL 1558  
Db 1818 EQARGQVTFV-GRRLRV-----HNRNLLSLDPDRVTRTEKIYDDHRKFTLRILYDQAGR 1871  
QY 1559 AFTNNIAPLDIVPEFKAKDGRVLGKIKQTLVKRVNYPN-----DILFLR 1604  
Db 1872 SLWSPSSRLNGVNTYSPGGYTAG--IORGIMSERMEYDQAGRITSRIIFADGKTWSTYILE 1930  
QY 1605 ET-----HSGAQMOLGVYRIRLNTLLASQLVSRAVNTGIDTILTMETORLPEPPIGEGFF 1659  
Db 1931 KSMVLLLSHQRY-----IFEPKNDRLSS--VTMPNVARQTLTIRS-----VGYY 1975  
QY 1660 ANFVLPKYDPABHG-----DERWPKIHIGNVGNVGRQPYYS-GMLS-----DTSE 1704  
Db 1976 RNI-----YQPPENASVIOQDFTEDGHLHTFVLG--TGRRVYIKYKGLSKLAETLYDTTK 2029  
QY 1705 TSMTLFVPAEYGM-----HEGVLGVGQKI-----TY 1734  
Db 2030 VSFT-----YDETAGMLKTNLQNEGFTCTIRYRQIGPLIDROIFFRTBEGMWNARFDNY 2085  
QY 1735 DMTWESAFYFDETRKQOQVFLINDA-----DHDSGMTQO-----G 1768  
Db 2086 DNS-----FVTSNQAV-INETPLPIDLYRDDVSGKTEQFGKFGVIYDIINOIITA 2137  
QY 1769 IVKNIKKYKGLNLSIATGYSAPMDFNSASALYYWELFYTPM-MCFORLLQEQOFDEAT 1827

Db 2138 VMTTKHFDAY-----GRMKEVOYEIERSLMTYMTVOYDNMGVRVKELKVGVYANTT 2190  
QY 1828 QWINVVVPAGYIVNGEIAPIWNCRPLEETTSW-----NANPLDAIDPDAVAQNDPMH 1881  
Db 2191 RY-SYEYDADQOLQTVSI-----NDKPL-----WRYSYDLNGN-LHLLSPGNSARLTPLR 2238  
QY 1882 YKIATFMRLLDQILIRGDMAYRELTRDALNAKMYVTRLELLGDEPDYSGQWAAASL 1941  
Db 2239 YDIR-----DRITRLGDOVK-MDEDF-----LRQGGDIFEYNSAGLLIKAY 2281  
QY 1942 SGMASTQVQAAAYQODLTMLGRGVSKNLTANSVLGLFLPEYNPA-----LTD 1989  
Db 2282 NRAGWSVRYXYDG-----LGR-RVSSKSSHHLQFFYADLTNPTKVTHLYNHSSEITS 2336  
QY 1990 YWQTLRLRLNLRHN-----LSID--COPLSL-----AIYABPT-- 2021  
Db 2337 LYDLOGLHFLAMELSSGDBFYIACDNIGTFLAVSGTGLMIKQILYATAYGEIYMDTNPF 2396  
QY 2022 -----DPKALLTSMVOASQ-----GGSAVLPGTSLYR 2049  
Db 2397 QIIIGYHGLYDP-----LTKLVHMGRRDYDVLAGRMTSPDHLWLKHLSSSNVMPFLYMPK 2453  
QY 2050 FPMVLERTRNLVAQLTQFGTSLLSWA-----EHDDADELTTLQLQGMELATQSI 2099  
Db 2454 NNPISNSQDIKCFMTDVNSWLLTFGFLHNVIPGYPKPDM-----AMEPSYELI 2504  
QY 2100 RIQORTVDEVDADI AVL-----ABSRRAQNRLEKYQOLYDEIDINHGRAMSLLDAAAG 2154  
Db 2505 HTQMKT-QEWDNSKSLGVQCEVQKQKAFVTLERFDQLYGSTITSCQAPKTKFASG 2563  
QY 2155 QSLAGQVLSIAEGVADLVPNVFLGACGSRGWAALRASAVM-----SLSATASQY-- 2205  
Db 2564 -SVFGKGVKPAKDGRTVTDIISVANEDGREAAILNHAHYLENLHFTIDGVDTHTYFKP 2622  
QY 2206 ---SADKISRSEAYRRRQWEIQRDNADGVKQMDAQLSILKIRREAAQO-----V 2255  
Db 2623 GPSEGDIALLGLSGGRRTLE-----NGNVTVSQINTVINGRTRRYTDIOQYGALCLNT 2677  
QY 2256 EYQETQQAHTQAQLELQKFTNKA-----LYSWMRGK-----LSAIYYQFF 2297  
Db 2678 RYGTTLDEKARVLELARQAVRQAWAREQQRLREGESGLRAWTEGEBKQQLVSTGRVQGY 2737  
QY 2298 DLTSQFCLMAQELRLRLTD--NGVTFIR 2324  
Db 2738 D---GFFVISVEQY-PELSDSANNIHFMR 2762

## RESULT 15

US-11-052-554A-171

; Sequence 171, Application US/11052554A

; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 171

; LENGTH: 2340

; TYPE: PRT

; ORGANISM: Rickettsia prowazekii

US-11-052-554A-171

Query Match 1.3%; Score 165; DB 7; Length 2340;

Best Local Similarity 18.1%; Pred. No. 0.052;



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 32.9071 Seconds  
(without alignments)  
4309.818 Million cell updates/sec

Title: US-10-754-115-45  
Perfect score: 7901  
Sequence: 1 MNSQDFSIETSLPKGGGA.....WFTVNEBNDTAAEVKKVKM 1474

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3569.5	45.2	1496	2 AH0447	insecticidal toxin
2	839	10.6	591	1 S09498	virulence-associat
3	835	10.6	591	1 S22664	virulence-associat
4	822.5	10.4	593	2 S15215	virulence-associat
5	709.5	9.0	563	2 S78224	virulence-associat
6	172.5	2.2	1404	2 E85509	hypothetical prote
7	172.5	2.2	1404	2 E90658	RhsG core protein
8	164.5	2.1	1399	2 A99720	RhsC core protein
9	161	2.0	1863	2 S46217	protein-tyrosine-p
10	160.5	2.0	1426	2 H64780	RhsD protein precu
11	158	2.0	3083	2 AH2493	hypothetical prote
12	156	2.0	2167	2 AF1489	cell wall-associat
13	155.5	2.0	1400	2 E90886	RhsG core protein
14	153	1.9	3036	2 T18995	hypothetical prote
15	151	1.9	1377	2 E86034	RhsA protein in rh
16	151	1.9	1394	2 H91236	RhsH core protein
17	151	1.9	1409	2 F91187	RhsA core protein
18	151	1.9	2314	2 T28698	hypothetical prote
19	149	1.9	1397	2 A85570	RhsC protein in rh
20	149	1.9	1398	2 B85549	hypothetical prote
21	149	1.9	1907	2 S50893	protein-tyrosine-p
22	148	1.9	1398	2 H90698	RhsD core protein
23	148	1.9	3972	2 S75251	hypothetical prote
24	147.5	1.9	682	2 C64898	RhsG protein - Bsc
25	147.5	1.9	4199	2 S76412	hypothetical prote
26	145.5	1.8	632	2 F83387	copper resistance
27	144	1.8	709	2 S38241	hypothetical prote
28	144	1.8	1377	2 C65159	RhsA protein precu
29	144	1.8	1411	2 E65145	RhsB protein precu

30	142.5	1.8	1512	2 AH0439	probable membrane
31	141	1.8	1274	2 T37193	enamelin matrix pr
32	141	1.8	1397	2 C64805	RhsC protein precu
33	141	1.8	2893	2 A64556	toxin-like outer m
34	139.5	1.8	1577	2 T30858	glucosyltransferas
35	139	1.8	828	2 G87584	hypothetical prote
36	139	1.8	2515	2 S47008	tenascin-like prote
37	138.5	1.8	1158	2 F90854	probable host spec
38	138	1.7	4427	2 P80637	polyketide synthas
39	137.5	1.7	656	2 G85731	Rhs element associ
40	137	1.7	2514	2 F81045	hemagglutinin/hemo
41	137	1.7	2628	2 T28651	hemagglutinin A -
42	136.5	1.7	1645	2 H85554	hypothetical prote
43	136.5	1.7	3566	1 A40701	tenascin-X precurs
44	136	1.7	775	2 C95921	hypothetical prote
45	136	1.7	1083	2 T23031	hypothetical prote

ALIGNMENTS

RESULT 1

AH0447  
Insecticidal toxin complex [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AH0447  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0447  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1496 <KUR>  
A:Cross-references: UNIPROT:Q8ZAV4; UNIPARC:UPI000000CDA67; GB:AL590842; PIDN:CAC93148.1.  
C:Genetics:  
A:Gene: YP03678

Query Match	45.2%	Score	3569.5	DB 2	Length	1496			
Best Local Similarity	49.3%	Pred. No.	3.8e-220						
Matches	748	Conservative	211	Mismatches	486	Indels	71	Gaps	29
1	MNS-QDPSITSLPKGGGAI	TGMEAL	TGPDGMAALSLPLPISAGRGYAPFTLNY	59					
1	MENSKQVAVAPLSLPKGGGAI	TGNGDSLGPISGMATLTLPLPISAGRGVAPSLTSLY	60						
60	NSGAGNSPFLGWDNCNVMT	IRRRTHFGVPHYDETD	TFLGPGSEVLVADQPRD--ESTLQ	117					
61	SSGSGNGPFLGWOLGTMA	IRRTNAQVPRYDEYDE	FLAPNGEVNVAADPGS	IERTEQ	120				
118	GINLGATFTVGYSRSLRSH	ESRLEYWPKT-TGKTDF	WLIYSPDGVHLLGKSPQARS	176					
121	SLN-GEQFSVIRYPRIEGN	FHRIEYMRPTNNSQAP	FWLHSSDQKHGCLYSGAA	RIA	179				
177	NPSOTTQTAOMLLASVSS	RGEQIYQVRAEDDTGCE	ADEI-----THHQA	QRYLHI	231				
180	DPLHPEHIAEWLLRESVLS	GEHIGYQADEQIDEP	SIYKAEKQNHPPASAQRYL	KR	239				
232	VYGNRTASETLPLGDSAP	QADWLFYLVFYDYGERS	NNLKTPTTAFSTTGSNLC	RDRFS	291				
240	VYGNRQAAYELCYLT-Q	QPAPTSLFSLIFDHGEYS	NIAEQVPVILKGSWNFRQ	DAFS	298				
292	RYEYGFETRRLCRQVLM	YHHLQALDSKITEHNGP	TVLSRLILNYDESATSLV	FVRR	351				
299	HFNTEYGFETRRLCRQV	LMYHNLKALGDBDPAQ	-TLVSRRLHYQHDATQ	LVGCCQ	357				
352	VGHEDQGNVVTLPPEL	AYQDFSPRHHAWPMDV	LANFNIAIQRMQLVDLKG	EGPLGLLY	411				
358	LAHPPDGTKRSLPPLPE	FDYQDFSTRDALGWQ	PLTDWAEFN--YQYQW	VDLNGEGNPGMLY	415				



A;Molecule type: DNA  
A;Residues: 1-593 <KRA>  
A;Cross-references: UNIPROT:P24419; UNIPARC:UPI00001780F6; EMBL:X56727; NID:G47836; PIDN:  
A;Experimental source: pEX102  
A;Note: the authors translated the codon CTG for residues 54 and 56 as Val  
R;Taira, S.; Baumann, M.; Riikonen, P.; Sukupolvi, S.; Rhen, M.  
FEMS Microbiol. Lett. 77, 319-324, 1991  
A;Title: Amino-terminal sequence analysis of four plasmid-encoded virulence-associated p  
A;Reference number: A54540  
A;Accession: A54540  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 1-10 <TAI>  
A;Cross-references: UNIPARC:UPI00001780F7  
R;Taira, S.; Rhen, M.

A>Title: Identification and Genetic analysis of mkaA-a gene of the Salmonella typhimurium  
 A:Reference number: S26565; MUID:90136009; PMID:2693884  
 A:Accession: S26565  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-56, 'P', 58-304, 'L', 306-340, 'R', 342-373, 376-492, 'V', 494-593 <TA2>  
 A:Cross-references: UNIPARC:UPI00001780F8; EMBL:Z15042; NID:g47782; PIDN:CAA78760.1; PID  
 A:Note: the authors translated the codon GTT for residue 493 as Gly  
 C:Genetics:  
 A:Gene: vscD; mkaA  
 A:Genome: plasmid  
 C:Superfamily: virulence-associated protein spvB  
 C:Keywords: DNA binding; transcription regulation

Query Match	10.4%;	Score	822.5;	DB 2;	Length	593;			
Best Local Similarity	47.3%;	Pred.	No.1.le-44;						
Matches	175;	Conservative	52;	Mismatches	116;	Indels	27;	Gaps	8;

  

Qy	9	ITELSLPKGGCAITGMGALPTPTGDPGMAALSLPLPISAGRGYAPAFPTLNVNSGAGNSPF	68
Db	15	ITPFPLPKG-----GKALSQSGFDGLASITLPLPISAERGFAPALALHYSSGGNGPFF	67
Qy	69	GLGWDGNVMTIRRRTHFGVPHYDETTDFLGPSGEVLV----VADQPRDESTLQ--GINLG	122
Db	68	GVGWSCATMSIARRTSHGVPOYNDSEDFLGPDPGEVLVQTLSTGDAPNPVTCFAYGDVSFP	127
Qy	123	ATPTVTGVRSLGSHFSLEYWQPTTKTGDTPWLIYSPDQGVHLGKSPQARISNPSTT	182
Db	128	QSYTVTRYQPTESFYRLYVGVNSNG--DDFWLLHDSNGILHLGKTAARLSDPQAS	186
Qy	183	QTAQWLKLEASVSSRGEQYIYQYRAEDDDTCEADEITHLQATAQRYLHIYYVGNRTASET	242
Db	187	HTAQWLVEESTPAGEHIYYSIAENGDNVDLNGEAGRDRSAMRYLSKVQYGNATPAAD	246
Qy	243	LPGLDGGAPSQADMFLFYLVFDYGERSNLNKTPPASTTGSWLRCRDRFSPSYEYGFETR	302
Db	247	L-YLMTSATPAQVQLFTLVVFEYGERGVDPQVPPAFTQAQNSWLARQDPFSLVNYGFEIRLH	305
Qy	303	RLCRQVLMYHLQALDLSKITEHNGPTLVSRLLINYESAIASTLVFVRRVHGEOQG----	358
Db	306	RLCRQVLMFHH---FPDELGE---ADTLVSRLLLEYDENPILTQLCARTLAYEGDGYRRA	360
Qy	359	---NVVTLPP	365

RESULT 5  
S78224  
virulence-associated protein mkfB - *Salmonella typhimurium* plasmid  
C:Species: *Salmonella typhimurium*  
C:Date: 29-Jan-1998 #sequence\_revision 13-Feb-1998 #text\_change 20-Aug-1999  
C:Accession: S78224; S23713; A43996  
R:Norel, F.  
submitted to the EMBL Data Library, January 1992  
A:Reference number: S78224  
A:Accession: S78224

A:Molecule type: DNA  
A:Residues: 1-563 <NOR>  
A:Cross-references: UNIPARC:UPI000000B8A96; EMBL:X57096; NID:g46998; PIDN:CAA40380.1; PID  
R:Norel, F.; Pisano, M.R.; Nicoli, J.; Popoff, M.Y.  
Res. Microbiol. 140, 455-457, 1989  
A:Title: Nucleotide sequence of the plasmid-borne virulence gene mkeB from Salmonella ty  
A:Reference number: A43996; MUID:90161559; PMID:2696057  
A:Accession: S23713  
A:Molecule type: DNA  
A:Residues: 1-513, 'NTYLEKGKGRILGDVAHFKEAEMLF', 514-563 <NOW>  
A:Cross-references: UNIPARC:UPI00001780F5; EMBL:X57096  
C:Genetics:  
A:Genome: Plasmid  
C:Superfamily: virulence-associated protein spvB  
C:Keywords: DNA binding; transcription regulation

Query Match 9.0%; Score 709.5; DB 2; Length 563;  
Best Local Similarity 39.1%; Pred. No. 1.7e-37;  
Matches 170; Conservative 60; Mismatches 138; Indels 67; Gaps 13;

Qy 9 ITLSLPKGGGAIATGMEALPTPTGDMGMAALSPLIPISAGRYAPAPATLNVNSGAGNSPF 68  
Db 15 ITPPFLPKG-----GKALSQSGPDGLASITLPLPISABERGAPA----- 53

Qy 69 GLGWDQNV-----MTIRRTHEG-----VPHVDETDITFLGPEGEVLV-----VAD 108  
Db 54 ---WRCITAAVAAMALRRGLVIRNEHCPPHQPVCVPQVINDSDEFLGPDGEVLVQTLSTGD 110

Qy 109 QPRDETLQ--GINLGATFTVTGYRSRLSHFSLEYWQPKTKTKTDFWLIYSPDGQVHL 166  
Db 111 APNPVTSFAYGDVSPQSYVTVRQPTRESSFYLEYWVGSNSG--DDFWLLHDSNGILHL 169

Qy 167 LKSPQARISNPSTOTTAQWLLASVSSRGEIYYQYRAEDDTCGCADEITHLQATAQ 226  
Db 170 LGKTAARLSDPQAASHATQWLLVESVTPAGEHIYYSYLAENGDNVDLNGNEAGRDSAM 229

Qy 227 RYLHIVYGNRTASSETLPLGDSAPSQADMLFVLVDYDYGERSNNLKTTPPAFTSTGMLCR 286  
Db 230 RYLSKVQIGNATPAADL--YLWTSATPAVQWMLTFLVDFGERGVDPQVPPAFTAQNSWLAR 289

Qy 287 QDRFSRYEYGEIETRIELCRQVLMYHLQALDSKITHENGFTLVSLRLNYSALIASTL 346  
Db 289 QDPFSLNYGPEIRLRLCRQVLMFHH---FPDELGE--ADTLVSRLLLEYDENPILTQL 343

Qy 347 VFERVHGQDQ-----NVVTLPPLELAYQDPSRHHAWQPMVLANFNAI-----Q 394  
Db 344 CAARTLAYEGDGYRRAPVNNMPPPPPPMMGGNSRRPKSKWAIWESKQIALRYTSAQ 403

Qy 395 RWQLVD--LKCEGLP 407  
Db 404 GYSVINKYLGGDYP 418

RESULT 6  
E85509  
hypoetical protein 20268 [imported] - Escherichia coli (strain O157:H7, substrain EDL5  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: E85509  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouasis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: E85509  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1404 <STO>  
A:Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI000000D025B; GB:AE005174; NID:g12512977; F  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z0268  
C:Superfamily: rhaF protein

Query Match 2.2%; Score 172.5; DB 2; Length 1404;  
Best Local Similarity 19.4%; Pred. No. 0.021;  
Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

Qy 327 PTLVSRILNLYDSAIASITLVFVRVVGHEQGNVVTLPPLLEYAYQDFSPRHH----- 379  
Db 51 PVLGAKVLPGETDLALPGPLPFI-----LSRAYSSVTRTPAPVGVFGP 94

Qy 380 HWQ--PMDVLANFNAINQWQLVDLKGBGLPGLLVQDKGAMWYRSQAORLGETGSDAVTWKEM 438  
Db 95 GWKAPFDI-----RLQIRD-----EGLILNDN-----GGRSIHFEPL 126

Qy 439 QPLSVIPSLQSNASLVNDINGDGLDQVITGFLRGVHSGRP--DGSWTRFTPLNALPVEY-- 496  
Db 127 FPGEI--SYSRSESF-----WLARG-GVAEQSSQLSALM-----QVLPEDEV 167

Qy 497 THPRAQLA--DLMG-----AGLSDLVITGPKSVRLYANTRDGF-----AK 534  
Db 168 LSPHMYLATNSLOQPPWILNWPVPGADEVLPEPPAYRVLTCGVVDGFGRTLAFRAAE 227

Qy 535 GKDVVQSGDITLVPVGADPRKLVAFSDVLGSGQHLVEV-----SATKVT--CWNPLGRGR 588  
Db 228 G-DV--AGAVTGVTDGAGR--FHLVL--TTOQRAEVFRKQKRSATSLSPAGRSASS 279

Qy 589 FGOPITLPGFSQPAEP-----NPAQVYLADLQSGPTDLIYVHTNRLDI 633  
Db 280 LVFPDTLPA-----GTEYGAQNGIRLEAWLTHDPA--YPDELPAAPLARYTYTASGELA 333

Qy 634 FLNKSQ--NGFABPVTLRFPEGLRFDHTCQLQWADVQGLGVASLILSVPHMSPH----- 686  
Db 334 VYDRSGTVQRGEA-----YDAEHAG-----RMVAHVAGR 363

Qy 687 ---WRCDLTNMFWLLNENNMVGHVHTLYRSSSOFWLDEKAAALTTGTPCYVLPFP 742  
Db 364 PESRYRYDDTGR---VTELNVPEGLDYREYGO-----DRVITDSLNRRELY--- 409

Qy 743 IHTLWQETDEISGNKLVTLRYARGAWDRERFRGFGVQETD-----SHOLAQ 794  
Db 410 -----TSGE--GGLKRVVKEHADGSIITRSEVDEAG--RLKAQTDAGRRTESLHWAS 459

Qy 795 GNAPERTPP-----ALTKWYATGLPVIDNALSTEYWRDQAPAGSPRFTT 841  
Db 460 GAVTAVTGPDRTVRYGYNQRQVTSVTPDGL-----RSSREY----- 498

Qy 842 WQNKQVPLTPEDNSRYWFRNALKGQLLRSLEYLGLDDSTNKHVPYVTFERSQVRLOH 901  
Db 499 -----DEKGRLAASERSGETTR---YSYDDPASE-LPTGIQD-----A 533

Qy 902 TDSRYPVLWSSVWESRNYHYERIASDPQSONIT--LSSDRFGQLKQLSVQYPRRQOPAI 960  
Db 534 TGSTKQWANS-----RYGQLLTFTDCSGYTRYEYDRYGO---QIAY-----HREGEI 578

Qy 961 NLYPDTLPDKLLANSYDDQORQLRLTYQQSSWHHLTNVTVRVVLGLPDSRSDI--FTYG- 1017  
Db 579 STSYNPNRQLVSKDAQGRETR--YEYSAAGDLT-----AIVAPDGRSEIYQDAMGK 631

Qy 1018 AENVPAAGL-----NLELLSDKNLIADDK--PREVLGOOKTAYTDGON----- 1059  
Db 632 AVSTTQGGLTRSMGYDAAGRITVLTNENGSQSTFRYPVDRLTQERG--PDGRTQRYHYD 689

Qy 1060 -TTPLOTPTQALIAFTTETTVFNOSTLSAFNGSIPSDKLSLTLTQAGYQOQNTYL--FPRT 1116  
Db 690 LTGKLTQSEDEGLVTLWHYDASDRITHRTVNGD-----PABQWYDEHGLWTLTSH 741

Qy 1117 GEDKVVVAHGYTDY-----TAAQWRRFQ-----KQSTQLTGK 1151  
Db 742 SEGRHVSVHYGYDDKGLRTGERQTVENPETGEMLWEHETGHAYSEQGLATRQEPGLPPV 801

Qy 1152 ITLIWDANYC-----VVVQTRD-----AAGLTSAKYDWRFLTP 1185  
Db 802 EWLTYGSGYLAGMKLGTGPLVYMRDRLHRETARSGEAYELATAWNTSGLRSRLNL 861



Qy	1186	VOLT-----DINDNQHLITLDALGRPITLRFWGTENGKMTGYSSEKASFSPPSDVNAIEL	1242
Db	862	PQLDRDWDNDNGOLIRISGPQESREYR--SDTGRLTGVHT-----TANLDIDI	910
Qy	1243	KKPLPVAQCQVYAPESWMPVLISQKTENRLAEQDWKLYNARIITEDGRICTLAYRRVQS	1302
Db	911	-----PYATDPAGNRLPD-----	923
Qy	1303	QKAIPOLISLNNQPRLPPLSHLTLTDDYDHPQIQIRQQ---VVFSDGFGRLLOAARH	1359
Db	924	-----PELHPDS-TLTA-----WPDNRITAEADAHVYRYDEYGRLEAKTDRI	963
Qy	1360	EAGMARORNEGDSLIINQVHTENRWAVTGRTEYDNKQPIRTYQPYFLNDWRVYNSDSAR	1419
Db	964	PEGVIRMHDE-----RTHHYHDSQHLRVFHTRIQ	993
Qy	1420	QEKEAYADTHVYDPIGRKIVITAKG-WFRRTLTFTPWFTVNEDENDT	1465
Db	994	HGEPQVESRYLYDPLGR---TGKRVRRERDLTGWMSLSRKPET	1036

RESULT 7

E90658

Rhsg core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C;Accession: E90658

R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference number: A99629; PMID:11258796

A;Reference number: A99629; PMID:11258796

A;Accession: E90658

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1404 <N>

A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BA033660.1

A;Experimental source: strain O157:H7, substrain RMD 0509952

C;Genetics:

A;Gene: ECe0237

C;Superfamily: rtsP protein

Query Match

Best Local Similarity 19.4%; Score 172.5; DB 2; Length 1404;

Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

Qy	327	PTLVSRLLINVDSEASTASTLVFVRVCHGEQDGNVTVLPPELAYQDFSPRHA-----	379
Db	51	PVLGAKVLPGETDLALGPLFFI-----LSRAYSSYRTTTPAPGVGFGP	94
Qy	380	HWC-PMDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWWYSAQRLGEIGSDAVTWKMM	438
Db	95	GWKAPFDI-----RLQIRD-----EGLINDN-----GKSIHFELP	126
Qy	439	QPLSVIPSLQSNASLVNDINGDGLDWVITGFLRGYHSQRP-DGSWTRFTPLNALPVEY-	496
Db	127	FPGEI--SYRSSESP-----MLARG-GVAEQHSSQPLSALW-----QVLPEDVR	167
Qy	497	THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRDGP-----AK	534
Db	168	LSPHWVILATNSLOQPMWILNWPVPVGADEVPPEPPAVRVLTVGVVDGFGRTLAFHRAE	227
Qy	535	GKDVQSGDITLVPFGADPKLVAPSDVLGSGGAHLVEV-----SATKVT--CWPNLGRGR	588
Db	228	G-DV--AGAVTGVTDGAGR---FHLVL-TTQQRAEVFRKQATSLSSPAGRSASSS	279
Qy	589	FGQPIITLPGFSPATEP-----NPAQVYLADLDGSGFTDLIVYHTNRLDI	633
Db	280	LVFPDITLPA---GTEGADNGIRLEAVWLTHDPA--YPDELPAAPLARYTYTASGELRA	333
Qy	634	FLNKGSG---NGPAPVTLRPEGLRFDHTCOLQWADVQGLGVASLILSVPHMSPHH----	686
Db	334	YVDRSGTGVGGA-----YDAEHAG-----RMVAHYVAGR	363

A;Accession: A99720  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1399 <HAY>  
A;Cross-references: UNIPROT:Q8X9E8; UNIPARC:UPI00001653CB; GB:BA000007; PIDN:BAB34152.1;  
A;Experimental source: strain O157:H7, substrain RIMD 050952  
C;Genetics:  
A;Gene: ECs0729

Query Match 2.1%; Score 164.5; DB 2; Length 1399;  
Best Local Similarity 19.6%; Pred. No. 0.067;  
Matches 305; Conservative 166; Mismatches 570; Indels 513; Gaps 80;

QY 18 GGAITG-----MGEALTPTGPGMAALSLPLPISACRGVAPAPITLVNSGAGNSPGLG 71  
DB 41 GGVTSGHVPNPLGAKVLPGETD--IALPGPLPILSRITYSSYKT---KTPAPVGSGLPG 95

QY 72 WDCNV---MTIRRT-----HFGVPHYDETDTFLGPEGEVLVADQPRDESTL----- 116  
DB 96 WKMPADIRLQLRDNTLILSDNGGRSLYPE-HLFPGEDG-----YRSSESLMLVRGGVA 147

QY 117 ---QGINLGATFTVTGYSRLSHESFSLVWQPKTKTDFWLIYSPDGQVHLIG---KS 170  
DB 148 KLDEGHLAALWQALPEELRLSPH-----RYLATNSPQGPWMLLWGCERV 192

QY 171 PQARINSPSTQTQTAQWLLLEASVSSRGE-QIYYQYRAEDDTGCEADRIITHLQATAQRYL 229  
DB 193 PEADEVLPAFLPPYR--VLTLGLVDRFGRTQTFHREAAEFSG-----EITGVTGAGRHFR 246

QY 230 HIVYGNRTASE-----TLPGI-----DGSAPSQADWLFYLVFDYGER 267  
DB 247 LVLTTQARAEARQAISGGTESPAFPDITLPGYTEYGRDNGIRLSAVMLTH-----DPEY 302

QY 268 SNNLKTPP-----AFSTGSMUCRODRFSR-----VEYFEIRTRRLC-----RQVLMTHHL 314  
DB 303 PENLPAAPLVRYGWTPRGELAVVYDRSGKQVRSFTYDDKYRGRMVAHRHTGRPEIRYR- 361

QY 315 QALDSKITEHNGPTLVSRLLIYNDESALFLVPVR-VGHEQ-DGNVVTLPPELAYQD 372  
DB 362 -DSGRVTEQLNPAGLS-YTYQYBKDRITITDSLNRRVLHTQEGGLKRVVKE----- 414

QY 373 FSPRRHAWQPMVDLANFNFAIQRWLKGEGLPGLLYQDKGAWWYSRQAQLGEIGSDA 432  
DB 415 -----HADGSVTSQFQDAVGR-----LRAQ-----TDAAGRTTEYSFV 448

QY 433 VTWEKQPLSVIPSLQSNASLVNDINGOLDWVITGP-GL-----RGHSQRPDG 481  
DB 449 VT-----GLITRITTPDGRASAFYNNHSSQLT-SATGPDGLFIRREYDEWGLRIOETAPDG 503

QY 482 SWTRFTPLNALPVEYTHPRAQL-----ADLMGAGLSDLVLGPKSVRLYVANTR-DGPAKGK 536  
DB 504 DITRY-----RYDNPHSLDPCATBDATG---SRKMTWVSRYGQLLSFTDCSGYVTRY 552

QY 537 DVVQSGDITL-----VPGADPR-KLVAFSDVLGSGQAHLVEVSATKVTCPNPLGRG 587  
DB 553 DHDRFGQWTAHREGLSQYRAYDSRGOLIAVKDTQGHETRYEYNAAGDLTT----- 604

QY 588 RFGQPIITLPGSPQATFENPAQVYLABLDGSGPTDLIYVHTNRDLIFLNKSN--GFAEP 645  
DB 605 -----VIAPDGRNGTQYDAMGKAICTTQGGTLRSMEYDAAGRIVRLTSENGSHTTFRYD 659

QY 646 VTLRFPESGLRFDHTCOLQADQVGLGVASLILSVPHMSPHHWRC-----LTN-----M 694  
DB 660 VLRLIQTGTGDTQRYHHDLTG-----KLIRSDDEGLVTHWHYDEADRLHRTVKGETA 715

QY 695 KPWLLNEMNNMNVHHTLR-YRSSSQFWLDEKAAALTTGQTPVCYLPFPPIHTLWQETED 753  
DB 716 ERWQYDERGMLTDISHISEGRVTVHYGYDEKGR-LTGERQTVHHQTEALLWQETRH 773

QY 754 EISGNKL-----VTLIRYARGAWDG-----REREFRGFGYVEQ 786  
DB 774 AYNAQGLANRCPDLSFAVEWLTYSGWLAKMGLDTPLVDFTRDLRLHRTLRFRGYEL 833

787 TDSHQLAQGNAPERTPPALTKNWYATGLPIDNALSTEYWRDQAFAGFSFRFTTWDNKK 846  
DB 834 TTAY-----TPAGQLQSQHLSNL-----QYDRD-----YTNDN- 862

QY 847 DVPLTPEDDNSRVWFRNALKGQLLR-----SELYGLDDS--TNKH-----VPYT 888  
DB 863 -----GELIRISSPRQTSYSYSDSGLRTGVHTTAANLDIRIPYA 902

QY 889 VTEFRSQVRELQ-HTDSRYFVLASSVVESSNYHYERIASDPQCSQNTILSSDRFGQLKQ 947  
DB 903 TDPAGNKLDPDELHPDSTLS--MMPDNRIARDAHY-----LYRYDRHGRLETEK 948

QY 948 LSVQYPRRQOPAINLYPDTLDPKLLANSYDQDQRLTTYQSSWHHLTNNT----- 999  
DB 949 -----TDLIPEGVIRT--DDETRH--YHYDSQHLVHVTRTQVEEPL 987

QY 1000 --VRVLGLPOSTSDIFTTYGAENVPAGLNLLEL-----SDKNSLTADDKPREYLG 1048  
DB 988 VESRYLYDPLGRRVAKRVRRERDLTGWMSLSRKPQVTWYGWDGDRLLTTIONDR- 1043

QY 1049 QOKTAYTDGONTTPTQTRQALIAFTETTVFNQSTLSAFNGSIPSKLSTLLEQAGVQ- 1107  
DB 1044 -IQTIYQGSFTPLIRVETATGELAKTORR-----SLADALQOOSGED 1085

QY 1108 -----QTNLFPRTGED-KVWVA----- 1124  
DB 1086 GGSVVPPVLVQMLDLRLESEILADRVSSESRMLASCLGLTVAQMQSQMDPVYTPARKIHL 1145

QY 1125 ---HHG-----YTDYGTAAQFWRP-QKQSNLTQLG-----KITLIWD 1157  
DB 1146 YHCDHRLPLALISKEGATEWCAYEDWGNLLBENPHQLQQLIRLPGQYDEESGLYNN 1205

QY 1158 AN-YCVVVQ---TRDAAGLTTSKDYDRF---LTPVQLTDINDNQHLITLDALGRPI 1208  
DB 1206 RHRYPDPLQGRYITQDDPIGL---KGNWNYQYPLNPVQY-----IDSMG--LA 1248

QY 1209 LRFWGTENGWMTGYSS--PEKASFPSPSDV-NAAIELKKPLPVAQCQVYAPESMMPVLQSK 1266  
DB 1249 SKYGHLLNG--GYGARENKPTPDPSKFPDIAKQLRLPYIDQA-----SSAPNVFKT 1299

QY 1267 TFRNLABQWKLQYNARIITFEDGRICITLAYRRVQSOKAIPOLISLNNGRPLP 1320  
DB 1300 FFRALSPDY-----TLYCRKWKV-----PNLTCTCPQDDPQYP 1332

RESULT 9  
S46217  
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat  
N;Alternate names: leukocyte common antigen-related phosphatase  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 07-May-1995 #sequence revision 03-Nov-1995 #text\_change 09-Jul-2004  
C;Accession: S46217; S51174; A49104  
R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.  
Biochem. J. 302, 39-47, 1994  
A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase  
A;Reference number: S46216; MUID:94347119; PMID:8068021  
A;Accession: S46217  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1863 <ZHA>  
A;Cross-references: UNIPROT:Q64605; UNIPARC:UPI000017705B; EMBL:L11587  
R;Goldstein, B.J.  
submitted to the EMBL Data Library, February 1993  
A;Reference number: S51174  
A;Accession: S51174  
A;Molecule type: mRNA  
A;Residues: 1-1788, 'G', 1790-1863 <GOL>  
A;Cross-references: UNIPARC:UPI000000EAD4; EMBL:L11587; NID:G205134; PIDN:AAC37656.1; PI  
R;Yan, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossie, K.; Musacchio, J.M.; Silvenn  
J. Biol. Chem. 268, 24880-24886, 1993  
A;Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the ner  
A;Reference number: A49104; MUID:94043351; PMID:8227050  
A;Accession: A49104

10



QY 1421 EKEAYADTH-VYDPI-GREI 1438  
 Db 1210 ESGLYNNRHYVDPLQGRYI 1229

RESULT 11  
 AH2493  
 hypothetical protein all7128 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120al  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
 C:Accession: AH2493  
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AH2493  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-3083 <KUR>  
 A:Cross-references: UNIPROT:O8YL10; UNIPARC:UPI00000CEEDB; GB:BA000020; PIDN:BA078212.1;  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: all7128  
 A:Genome: plasmid

Query Match 2.0%; Score 158; DB 2; Length 3083;  
 Best Local Similarity 18.8%; Pred. No. 0.66;  
 Matches 276; Conservative 163; Mismatches 507; Indels 520; Gaps 73;

QY 1 MONSODFSITELSLPKGGAITMGCEAL-TPTGPDG-----MAALSILPLPISAGRYA 52  
 Db 725 IEAAEVNVAASAGAGGGIIGTGLNLKDPFGDKVGRNEFVQLLNPIEMPDASGLV 784

QY 53 PAPTLYNYSAGN-----SPFG-----LQWDCNVMTIR----- 80  
 Db 785 QAYLMAYAKVAGKVKRIESPVTLLPGYKVSETPPQLHLATDIGGNLRINMGPNAAA 844

QY 81 -----RRTHEGV-----PHYDSD 94  
 Db 845 REINTEDEAEVFTVTTDGKLTVSAFNIPTQYSGVSKIADGGTKNDTIEIKPIEISA 904

QY 95 TFLGPEGVLVADOPRESTLQINLGATFTVTGYRSLRSHFSLRYWQPKTKTKTDF 154  
 Db 905 DLKGAGEDLYGSGSD--TIRG--GAD-----WDLGYGDRDD 940

QY 155 WLIYSPDQGVHLLGKSPQARISNPSQTTQTAQWLLLEASVSRGEQIYYQYRAEDDTGCEA 214  
 Db 941 F-VYGDGDWLDG-GAGADILGCGAGPDT-----ASVTSATSAISINLVQVSTGDAA 992

QY 215 DEITHHL-QATAQRYLHV-----YGNRTASETLPG-----LDG--- 248  
 Db 993 DDVFQSIQIVGSRDYDTLIGDEDNNEFDGEGENDFISGGAGDDRLSPGWGDVDIDGTG 1052

QY 249 -----SAPSOA-----DWLFYLVDPYGERSNLKTTPPAFTSTGSMWLCRQD 288  
 Db 1053 TDTLVIDYSSLPQTQAVANSELDPNSTDFVYVANAIGICA-PIKTD--INVSGN----- 1103

QY 289 RFSRYEYGEINRRLCRQVLMYHHLQALDSKITEHNGFTLVSRILINLYDESATLTVF 348  
 Db 1104 -----YHATLSAD-----GLTVAGSGILGNGSGS--GNQGLW 1132

QY 349 VRRVCHQDGNVVTLPPELAYQDPSRHHAH--WQPMDV--LANFNALQWOLVDLK-- 402  
 Db 1133 VKKI-HSSDPAVRVTPNNQV-YQPLLSEDSKVVWSQGSSTWANTNGTQVRQLTKLSIN 1190

QY 403 -----GEGLPGLLYQDKGAWVYRSARLGEI-----GSDAVTWKWKQPLSVIPSLQSN 450  
 Db 1191 IGYGDGYLATISEDGSSTIAWLRKRNKFTYTFIANADGNLRQIN-----IFTGGG 1246

QY 451 ASLVDINGDG-QLDWVITGPGLRGYHSQRPDGWSWTRFTPLNALPVEYTHPRAQLADLMGA 509

Db 1247 VRELDLSADGSKITWSQDG-----GY-----GPGGVWVANT-----DGTNIRELSGNLGY 1292  
 QY 510 GLSDLVL-----IGPKSVRLYANTRDG--FAKGDVVQSGDITLTPVPGADPR 554  
 Db 1293 NINPSISADGSTVVWAGYQAGYASTNLAAATDGSRFVWVPNTTEEVGEPAQOSLAGDSR 1352

QY 555 KLV-----AFSDVLGSGQAHLEVEVSATKVCWPNLGRGRFGQPIITLPGFSQAT 603  
 Db 1353 RWFVTKFNGSDYSLVYGDIDGIBQILIDASS-----PNIGIGRGHALSVDLVGRYN 1406

QY 604 EFNPA-----QVYLADLDSGGPTDIYVHTNRDLDF-----LNKS--GNGPAEP 645  
 Db 1407 SFDPAUSGEIYT-----WGPSRIKYSNFERFDIIGTRYGDELFGNLDLSLMMGGGAD- 1460

QY 646 VTLRFPPEGLRDPDHTCQLQMDADVQGLVASLTLSPHMSPHHRCDLTNMKPWLNMENNN 705  
 Db 1461 -TLK--AGLGDD---IYILDTQAGGSQIE-----DAGGTD--TLRLTTRN 1498

QY 706 MGVHHTLYRSSQFWDKAAALTTGTQPCVYLPFPPIHTLWQTET-----EDEISGN 758  
 Db 1499 PGATNTPRITDA-----DLSLAVPTTG-----IFGMRRAGTSLIIDLNKDGIAAS 1543

QY 759 KL-VTTLAYARGANDGREBERFGRGYVEQTDHQLAQGNAPERTPPALTKNWYATGLPVI 817  
 Db 1544 KTDLTILNFFDTVTGTG-----AGTGFIETVANLAGAE-----ILSK-----LQVG 1583

QY 818 DNALS-----TEYWRDDQAFAGFSPRFTTW--QDNKDVPILTPEDNSRY----- 859  
 Db 1584 DDTISGSNAADFDIGWLSNDTLTSGAGNDTLWGQDGNDF--LNGEDGNDLSQGGNGNDTLT 1642

QY 860 --WPNRAKGLLRSELYGLDSTNKHVPYTVTFERSQVRRLQHTDTSRYPLVWSSVBSR 917  
 Db 1643 PGWGNVVVDGG-AGTDVLVLD-----YSLNLTNA-----VAVRTLSGTS 1680

QY 918 NYHYERTASDPCCQNITLSSDRFQPLKQLSVQYPRQQAIPALNYLPTDLPKLLANSYD 977  
 Db 1681 GNYLOKP-----FIGNAYGLGTPLKIRET-----NSVSDKXFLSA-- 1715

QY 978 DQORQLRLTYQQSSWHHLTNNTVRLGLPDSRSDIFTTYGAENVPPAGLNLELLSDKNSL 1037  
 Db 1716 -----DGTYYAVYTYVYNDPANGLIWIKKIDDSGL 1746

QY 1038 IADKPPEYLG-----QOKTAYTDGNTTPTLOTPTROALIAFTETTVFNQSTLSAFNGSIP 1093  
 Db 1747 VKIDEIATEIALSTDGKIAMSDG-----RVYVANTNGTEKIRINLNNINGIY 1796

QY 1094 SDKLS-----TTLEQAGYQOQNTYLPFRTGEDKVVVAHGYTDYG 1132  
 Db 1797 SLSLSGDSQSVWNGNQLLVANTDGTNIREITQSSTKSFISENGSQIHWAGYQG-EKYG 1855

QY 1133 -----TAAQFWRPQKQSNLTQTKG--ITLIWDANYCVVYVQTRDAAGLTTSK- YDWRF 1182  
 Db 1856 IWSASTSTSLPVKSLVDGNLSLSSDGIKAIWQDRYFLSVSSTNSTEQVAESYDFRV 1915

QY 1183 L---TPVOLTD-----IN-DNQ 1195  
 Db 1916 VGSSEPVLAADGAKVAFIKAINADNQ 1941

## RESULT 12

AP1489  
 cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - L  
 C:Species: Listeria innocua  
 C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AF1489  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker  
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fshih, H.  
 D.; Jones, L.M.; Karst, U.  
 Science 294, 849-852, 2001  
 A:Authors: Kref, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma  
 ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
 A:Title: Comparative genomics of Listeria species.



Db 276 LSDSAPPDTLPG-----TEYGPDRGIRLSAVWLTHDPAYPESLPAGPLARYTYTIBAGBL 330  
QY 633 IFINKS-----GNGFAEPVTLRFPEGLRPDHTCQLOMADVOGLGVAS 674  
Db 331 AVYDSNTQVRAFTYDAQHPGRWAHYRAGRPEMY-----RYDDTGRV-----VEQLNPAG 382  
QY 675 LILSVPHMSPHHRCOLTNMKPWLNNMNNMGVHHTLR--YRSSQFWLDEKAAALTTG 732  
Db 383 LSRYQYEQDRITVTDLSLRREVLHTE--GGAGLRKVVVKELADGSVTHSGYDAAGRLLTA 440  
QY 733 QTPVCVCLPPIHTLQTEDEISGNKLVTLRYARG-----AWDGREREB----- 778  
Db 441 QT-----DAAGRTEYGLNVVSGDITDITPDGRETKEFYNDGNOL 481  
QY 779 -----RGFG-----YVEQT-----DSHQLAQGNAPERTPPALTKN-----W 809  
Db 482 TAVVSPDGLSRRAYDEPGRLVSETSRCGDVIRYAYDN-PHSELPAITTDATGSTQMTW 540  
QY 810 --YATGLPVID--NALST--EYWRDQAFA-----GFSRPTTWQDNKQVPLTPED---DN 856  
Db 541 SRYGQLLAFTDCGYQTRYEDRFQGMTAVHREEGIS-RYRRY-DNRGRLTSVKDAQGHE 598  
QY 857 SRYWPNRA-----LKGQLLRSELYGL-----DGS 880  
Db 599 TRYETNAAGDLTAVITPDGNRSETQYDANGKAVSTTQGLTKSMYDLAGRITTLTNG 658  
QY 881 TNKHVPYVTEPRSVRRLOHTDSRYPLVMSVSRVNYHYERIASDPQCSQN--ITL-- 936  
Db 659 SRSEFTYDALDELVOORGF--DGR-----TQRYHYDLTKLQSEDEGLVTLWH 705  
QY 937 --SSDRF-----GQPLKQLSVQYPRR-----QQPAINLYPDTLPDKLLANSY 976  
Db 706 YDESRLTHRTWNGSPAEQW--QYDEHGWLTEISHLSEGHQVAVHY-----GY 751  
QY 977 DDOQRQL--RLTYQSS-----WHLTNNYTVRLGL-----PDS--TRSDIPTYGAEVPA 1023  
Db 752 DDKGLAGERQVHNPEGTCELLWQHETBHAYNEOGLANRVTPDLSRVEWLTYS----- 806  
QY 1024 GGLNELLSKNSLIADDKPREYLQOQKTAYTDGQNTTPTQPTTQALIAFTETTTFVFNQS 1083  
Db 807 -----CYLAGMKL-----GGTFLVEFTDRL-----HRE 830  
QY 1084 TLSAFNGSIPSKLSTLEQAGYQNTNLYFPRTGDKGVVAAHGYDYGTAQFWRPQK- 1142  
Db 831 TVRSPGNN--AVELTSTVTPAGHLQSORL-----NSQVVDYRDYDNDNGDLVRISGPRQT 883  
QY 1143 -QSNTQLTKI---TLIWANYCVVQTRDAAG-----LTTSA 1176  
Db 884 WEYGSATGRLESVETLASDLDIRIPYAT-DPAGNRLPDPELHPDSTLTAMPDNKRIADA 942  
QY 1177 KYDWRFLTPVQLTDIN-----DNQHLITL----- 1200  
Db 943 HVYRHDEYGLRTEKTDRI PAGVIRTDDBERTHHYHDSQHLRVFYTRIQHGEPLVESRYL 1002  
QY 1201 -DALGRPTTLRFWTEGKMTGYSS-----PEKASFPSPSDVNAALTELKPLPVAQCQ-VY 1254  
Db 1003 YDPLGRMAKRWRRER-DLTGWSLSRKPEVTWYGDGDLRTTVQ-----TDTTRIQTVY 1057  
QY 1255 APESHMPVLSQKTFNRLAEQDWQKLYNARIITEDEGRICTLAYRRWVQSKAIPQLISLN 1314  
Db 1058 BFGSFTPLIRVETENGEREKA--ORSLAETLQOQBS-----E 1093  
QY 1315 NGPRLP-PHSLTTLTDYDHDPEQOIQVVFSDGFGRLQAAARHEAGMARQORNE--G 1371  
Db 1094 NGHGVVFAELVRLDLR-----EEIRADRVSSESRAWLAQCGLTVEQ-LARQVEPEYTP 1148  
QY 1372 SLIINVQHTENR-----WAVTGRTEYDNKGQPIRTVQPYFLND-WRYVSND 1416  
Db 1149 ARKVHFYCHDRGLPLALISEDNTAW---RGEYDEWGNOLNEENPYLLQPVRLPQQ 1204  
QY 1417 SARQEKEAYADTHVVDPI-GREI---KVITAKGW 1446  
Db 1205 HDEESGLYNNRNYDPLQGRYITQDPIGLAGGW 1238

## RESULT 14

T18995  
hypothetical protein C06B8.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18995  
R:Steward, C.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19057  
A:Accession: T18995  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-3036 <WIL>  
A:Cross-references: UNIPROT:O17575; UNIPARC:UPI000017B740; EMBL:Z81463; PIDN:CAB03852.1,  
A:Experimental source: clone C06B8  
C:Genetics:  
A:Gene: CESP:C06B8.7  
A:Map position: 5  
A:Introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/3  
Query Match 1.9%; Score 153; DB 2; Length 3036;  
Best Local Similarity 18.7%; Pred. No. 1.3; Mismatches 459; Indels 428; Gaps 65;  
Matches 241; Conservative 164;  
QY 335 LNYDESATIASTLVFVRRYVGHEQDGNVVTLPPLLEYAQDFSPRRHHAHQWMDVLNFNATIQ 394  
Db 118 INYDSEMPNFRLV-----DGTPTVQRLQVQFRD-----RWRSVCTM-----VT 156  
QY 395 RWQLVDLKGEGLLGYODKGAW-WYR-----SAQRLGEI----- 428  
Db 157 NWTSID-TGTACRSNGYSDGGFWKFRNRNDTYPFVMPKPDCHGAANKLWDCPAFSNPOK 215  
QY 429 -----GSD-----AVTWEKQPL-SVIPSLQSNASLVDIN--GGQ 461  
Db 216 IRLSENLCQGEDDITGIYCWGPPPTTGWARHWKQILNSPFHYVNSDPDLVAVNRESDR 275  
QY 462 LDWV-ITGPGLRGYHSQRPDGSWTRFTP--LNALPVEYT-HPRAQLADLMGAGLSDLVLI 517  
Db 276 LEFVDILVAGYGVNKNVTSALYIEGVPPINNGRIEHSARDGLQLLDANGPA----- 328  
QY 518 GPKSVRLYANTRDGFAGKQVQSGDITLVPAGADPRKLVAFSDVLGSGQAHLEVSATK 577  
Db 329 -----ITANSTFSYNRGHGI-----SVVNTTDARIF-INNTK 359  
QY 578 VTCWPNLGRGFRGQPIITLPGFSQPAETHENPAQVVLADLD-----GSGPTDLIYVHTNR 630  
Db 360 I-----QGNWGDGI-----WYKOQTGVN-----LIDYGMRRERSIGSRLE---EQKPR 400  
QY 631 LDI FLNKSNGFABFPVTLRFPEGLRFDHTCLOMADVOGLGVAS-----LILSVPHMSPH 685  
Db 401 IDMCAEH-----RVDDNHFFPHLIAVNLKNRTYLDLAQPAICWMTVSLPRLPY 449  
QY 686 HWRCDLTNMKWLNNMNNMGVHHTLYRSSQFWLDEKAAALTTGQTPVCYLFPFPHYHT 745  
Db 450 -----TVSIQWLHIRDNPQTARTTLVACDSNN--VDENSCSTPRFPIRNEIFP--- 498  
QY 746 LMQTEDEISGNKLVTLRYARGAWDGREREPFRGFGYVEQTDHQLAQGNAPERTPPAL 805  
Db 499 ---QSILKSSGKPLYLALEH---VLDGDAQ-----GYV-QGDVHLLFNIIHA-----SVL 541  
QY 806 TKNYWA-----TGLEVIDN-----ALSTEYWRDDQAFAGFSFR----- 838  
Db 542 DKAYYGLNVNTNCIIKNTKNGVFANDIRERTALTNVTLDENQGYAGFLVKDGAADWLNE 601  
QY 839 ---FTTWQD-----NKDVPLTP-----EDDN 856  
Db 602 TRILNNMGDMNISYAGGSIWVNGTRIEKNRWRGAIIHYNQTLPLPMPYNEVIFKGRPSN 661  
QY 857 SRYWPNRAK---GQLLRSELYGLDDSTNKGHPVYVTFEFSQVRRLOHTTSRYPVL--- 909  
Db 662 NKFYLPITISENWEGLLVGNFCAYSNESNHSKTRIP-----INSVTPKGVPK 710



QY	910	----	WSSVSRNHYHERIAS--DPOCSQNTILSSDRFGQPLKQLSVQYPRRQOPAINLY	963	Db	125	HLFPGBDGY--SRSESLWLVGVGAKLDEGHLRALUQALPBEELRLSPHRYLATNSPQGP	182
Db	711	ILISWVFLKNQVHPSMEIFSCDPDVINWIV--DVTGNRIDG-NLGYGWRAPAVNMH	766	QY	506	--LMG----	AGLSDLVLIGP-KSVRLYANTRDGFAGKQDVVQ-----SGDITLPPVGAD	552
QY	964	PTLDPKILANSYDDOQRLRLTYQSSWHHLTNNTVRLVGLPDSRSDIFTYV-AENVP	1022	Db	183	WLLGWCEVPEADEVLPAFLPPYRVLTLGLVDRFGRTQTFHREAAGEFSGEITVTDGAG	242	
Db	767	TWINSNQFLHN-----DITLYIRNAQWPELGDLPV-----TISKNVFKNVAKYII	815	QY	553	PR-KLVAFSDVLGSGQAHVVEVS-ATKVTCPNPLGRFRGQPIITLPGFSQDPATEFNPAQV	610	
QY	1023	AGLNLELLSDKNSLIADDPREVL--GQ-----KTAYTDGQNTPTPQTQALIAFT	1075	Db	243	RHPRVLVTTOAQRAEERQAISGGTEPSAFPD-----TLPGYTE-----	282	
Db	816	SIGN-----EDAPROFLTNQOONEIRANTVDFPFPPLSPRSTPYAALVVSSS	863	QY	611	YLADLDSGPTDLIYVHTN--RLD-IFLANKSGNGFAEPVTLRPPEGL-----RFDHTCQ	661	
QY	1076	ETTVFNQSTLSAFNGSPSKLSTTLBOA-----GYQTNLYLPRTGEDKVVWA	1124	Db	283	-----YGRDNGIIRLSAVWLTH-----DP-----EYPENLPAAPLVRYGWTGR	319	
Db	864	NVKIHR-----NCFNERSKYEIAIELEQHAKWADARENWGFQ-----IPRF-IDKPF--	912	QY	662	LOMA-----DYQGLGVASLILSVPH-----MSPH-----HWRCDLTNKKMFWLLNEMNNMG	707	
QY	1125	HGTYDTGTAAQFWRPOKQNTQLTGKITLIWDANY-----CVVVQTRDAAGL	1172	Db	320	GELAVVYDRSGKQVRSFTYDDKYGRMVAHRHTRGRPEIRYRYSDCR-----VTEQLNPAG	375	
Db	913	DQFNRYSLASIDIEPYMAACNQRPYITLLNGAFRQKQSPFKLGGIYENHDLKKG	971	QY	708	VHHTLRY-----RSSQFWLDEKAAALTTGQTPVCYLPFPPIHTLMQTFETEISGNKLVTTL	764	
QY	1173	TTSKYDWRFELTPVQLT-----DINDNOHLITLDALGRPIITLRFWGTENGQWGYSP--	1225	Db	376	LSYTYQEKDRITITDSLNARREVLHTQGB-----GGLKRWVKK	413	
Db	972	RYQVTEDLQVVPKAKLTIASGVLFOHGIQMLVQGDILRNEY--DQDEKVIETSTPFTL	1029	QY	765	RYARGAWDGRERFRFGFYVE-QTDSHQLAQGNAPERTPPALTKWYATGLPIDNALST	823	
QY	1226	EKASFSPDVNAAIELK-----KPLVAQCVVAPESWMPVLSQKTFNRLA--SODWQKLY	1280	Db	414	EHADGS--VTQSFDAVGRLEAQT-----AAGRTTSEYSPDVVT--GLITRTITPPDGRASA	465	
Db	1030	ERRSDIBLVDVGNVYETGRLEVLVDQGTVCNRSWTFQLTILACNLGLVADVQYFE	1089	QY	824	EYWRDDQAQFAGSPRFTTWQDNKV-----PLTPEDDNRYWFNFRALKQGLLR	871	
QY	1281	NARIITEDGRICTLAYRRWVQSKAIPLQILSLANGPRLPHSLITLTTDRYDHDPEQOIR	1340	Db	466	FYNHHSQUTS-----ATGPDGLEIRREYDEWGRLLQETAPDGDITRYR-----	510	
Db	1090	NWRLFPEPGD-----LPMINDNIRCEENEVDITKCRH-----	1121	QY	872	SELYGLDDSTNKHVPYVTFEFSQVRRLOHTDSRYPLWSSVVSRYHYERIASDPQCS	931	
QY	1341	QQVVFSDGFRLLQAAARHEAGMARQNEGDGSLIINVQHTENRWA-----VTG	1388	Db	511	-----DNPHSLDLP-CATE-----DATSGKMTWS-----RYGQLLSFTDCS	546	
Db	1122	-----DGVERNCAAGCR-----STEVVGLRCLEPRWAGVRYSLLANPPTVTG	1163	QY	932	QNTI-LSSDRFGQ-----PLKQLSVQYPRRQOPAINLYPDTLPKLLANSYDQOQRQ	982	
QY	1389	RTEYDN-----KGQ--PIRT--YQPYFLNDWRY	1412	Db	547	GYVTRYDHRDFRGQMTAVHREEGLSQYRAYDSRGLLIQAVK-----DTQGHG	591	
Db	1164	QTTMDNWRIBKGLFNFTSEFAAFKVDWNY	1195	QY	983	LRLYQOSSWHHLTNNTVRLVGLPDSRSDIFTYGAENVPAAGLNLLELLSDKNLSIADDK	1042	
RESULT	15			Db	592	TRYENNA-----		
E86034				QY	1043	PREYLGQOKTAY-----TDGQNTPLQPTTQOALIAFTTTFVNOSTLSAFNGSIPSDK	1096	
rnsA protein in rns element [imported] - Escherichia coli (strain O157:H7, substrain EDL				Db	611	SRN--GTQYDAWGKAICTQGGGLTRSMEYDAAGRVRLTSE-----NGS-----	652	
C;Species: Escherichia coli				QY	1097	LSTTLQAGVQOQNYLFPRTGED-KVVAHGHGTYDTGTAAQFWRPOKQNTQLTGKITLI	1155	
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004				Db	653	-HTTFR---YVDLRLIQTQGTGFRQRYHH-----DLTGKL-----	685	
C;Accession: E86034				QY	1156	WDANVCVVVQTRDAAGLTTSKAYD-----WFLTPVQLTDINDNOHLI	1198	
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew				Db	686	-----IRSED-EGLVTHWHYDEADRLTHRTVKGETAERWYDGRGMUTDIS---HI-	732	
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,				QY	1199	TLDALGRPTLRFWGTENGKWTG-----YSSPEKASFPSPSDVNAAIELKK	1244	
Nature 409, 529-533, 2001				Db	733	-----SEGRVTVHYDEKGLTGERQTVHHQPTEALLWQHETRHAYNAQGLANRCI--PD	787	
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.				QY	1245	PLPVAQCVVAPESWM-----PV-----LSQKTFNRLAEQWQKLYNARIITEDG	1289	
A;Reference number: A85480; MUID:21074935; PMID:11206551				Db	788	SLPAVEWLYTG-SGWLAKMKLGDTPLVDFTDRDLRHKRLRRFRGYELTTAYTPAGQLQSQ	846	
A;Accession: E86034				QY	1290	RICTLAYRR---WVQSKAI-----POLISLNNNGPRLPHSLTLTTD---RYDHD--	1334	
A;Status: preliminary				Db	847	HMSLQYDRDYYTWNNGELIRISSPQTRSYSDSGRLTGVTHTTANLDIRIPYATDPA	906	
A;Molecule type: DNA				QY	1335	-----PEQTRQOVVVS-----DGFGRLLIQAARHEAGMARQNEGDGSL	1373	
A;Residues: 1-1377 <STO>				Db	907	GNRLPPELHPDSTLSWPNRRIARDAHYLYRVDHGRGLTEKTDLTLPPEGVIRTDDE---	962	
A;Cross-references: UNIPROT:Q8X418; UNIPARC:UPI000000DDCF; GB:AE005174; NID:gl2518335; E				QY	1374	IINVQHTENRWAVTGRTEYDNKQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVYDP	1433	
A;Experimental source: strain O157:H7, substrain EDL933				Db	963	-----RTHRYHDSQRLVHYTRTQYEEPLVESRYLYDP	99	



Search completed: February 16, 2006, 21:44:59  
Job time : 41.9071 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 197.739 Seconds  
(without alignments)  
5259.201 Million cell updates/sec

Title: US-10-754-115-45  
Perfect score: 7901  
Sequence: 1 MQNSQDPSITSLSPKGGGA.....WFTVNEBNDTAAEVKKVKM 1474

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	Q8GF99 PHOLU	Q8GF99 photorhabdu
2	7285.5	92.2	1475	Q7N7Y2 PHOLL	Q7N7Y2 photorhabdu
3	6065.5	76.8	1476	Q7N7Z0 PHOLL	Q7N7Z0 photorhabdu
4	6031.5	76.3	1476	Q93EP6 PHOLU	Q93EP6 photorhabdu
5	4627.5	58.6	1485	Q7N936 PHOLL	Q7N936 photorhabdu
6	4595.5	58.2	1485	Q85153 PHOLU	Q85153 photorhabdu
7	3984.5	50.4	1433	Q8AP56 YERFR	Q8AP56 yersinia fr
8	3971	50.3	1428	Q9F9Z2 9ENTR	Q9F9Z2 serratia en
9	3638	46.0	1489	Q6QDZ4 YERPS	Q6QDZ4 yersinia ps
10	3591	45.4	1481	Q693A4 YEREN	Q693A4 yersinia en
11	3569.5	45.2	1496	Q8ZAV4 YERPE	Q8ZAV4 yersinia pe
12	3569.5	45.2	1516	Q8D1PC YERPE	Q8D1PC yersinia pe
13	3350.5	42.4	1401	Q93RN8 XENNE	Q93RN8 xenorhabdu
14	2277.5	28.8	1447	Q4ZP56 PSESM	Q4ZP56 pseudomonas
15	2274	28.8	1446	Q87X46 PSESM	Q87X46 pseudomonas
16	1076	13.6	2439	Q4HWU2 GIBZE	Q4HWU2 gibberella
17	837	10.6	591	1 VBP2 SALCH	P17450 salmonella
18	835	10.6	591	1 VBP2 SALCH	P74846 salmonella
19	832	10.5	591	1 VRP2 SALTY	P21454 salmonella
20	829	10.5	591	1 VRP2 SALTY	P55220 salmonella
21	826.5	10.5	593	1 VRP2 SALDU	P24419 salmonella
22	826.5	10.5	593	1 VRP2 SALDU	Q4L1X7 salmonella
23	709.5	9.0	563	2 P72421 9ENTR	P72421 salmonella
24	676	8.6	283	2 Q8KSN2 XENNE	Q8KSN2 xenorhabdu
25	651.5	8.2	2031	2 Q63XE6 BURPS	Q63XE6 burkholderi
26	502	6.4	1806	2 Q62N48 BURMA	Q62N48 burkholderi
27	457.5	5.8	166	2 Q6WCC1 9ENTR	Q6WCC1 serratia pr
28	386	4.9	2217	2 Q8TP72 METAC	Q8TP72 methanosarc
29	330.5	4.2	2379	2 Q72U39 LEPIG	Q72U39 leptospira
30	326.5	4.1	221	2 Q58FC3 SALGI	Q58FC3 salmonella
31	292	3.7	2306	2 Q8F107 LEPIIN	Q8F107 leptospira

32	289.5	3.7	2554	2	Q72QR5 LEPIC	Q72qr5 leptospira
33	279.5	3.5	2321	2	Q8F5B9 LEPIN	Q8f5b9 leptospira
34	241.5	3.1	1826	2	Q987Z7 RHILLO	Q987z7 rhizobium 1
35	202	2.6	2370	2	Q82RE3 STRAW	Q82re3 streptomyce
36	199	2.5	2060	2	Q6QBO0 STALU	Q6qbo0 staphylococ
37	181.5	2.3	2364	2	Q82R58 STRAW	Q82r58 streptomyce
38	180	2.3	1976	2	Q8D4R5 VIBU	Q8d4r5 vibrio vuln
39	177	2.2	795	2	Q7NFS7 GLOVI	Q7nfs7 gloeobacter
40	174.5	2.2	1881	2	Q8TJS7 METAC	Q8tjs7 methanosarc
41	172.5	2.2	1404	2	Q7AH10 ECO57	Q7ah10 escherichia
42	172.5	2.2	1404	2	Q8XED9 ECO57	Q8xed9 escherichia
43	171.5	2.2	2566	2	Q8TSE7 METAC	Q8tse7 methanosarc
44	170	2.2	593	2	Q8GP37 PSEAE	Q8gp37 pseudomonas
45	170	2.2	2082	2	Q9ACP4 STRCO	Q9acp4 streptomyce

## ALIGNMENTS

RESULT 1  
Q8GF99 PHOLU PRELIMINARY; PRT; 1474 AA.  
AC Q8GF99  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE TcdB2.  
GN Name=tcdB2;  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=29488;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;  
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Ferry R.D.,  
RA french-Constant R.H.;  
RT "The tc genes of Photorhabdus: a growing family."  
RL Trends Microbiol. 9:185-191(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;  
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;  
RT "Genomic islands in Photorhabdus."  
RL Trends Microbiol. 10:541-545(2002).  
DR EMBL; AF346500; AA017202.1; -; Genomic\_DNA.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integrin complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR000413; Integrin alpha.  
DR InterPro; IPR003284; Sal\_SpVb.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF03534; SpvB; 1.  
DR PRINTS; PR01341; SALSPVBPROT.  
SQ SEQUENCE 1474 AA; 166338 MW; 9118D4E4914683DD CRC64;

Query Match	100.0%;	Score	7901;	DB 2;	Length	1474;			
Best Local Similarity	100.0%;	Pred. No.	0;	Mismatches	0;	Indels	0;	Gaps	0;
Matches 1474;	Conservative	0;							
Qy	1	MONSQDPSITSLPKGGGAI	TGMG	ALPTGPDG	MAALSPLPIS	AGRGVAPFTL	NN	60	
Db	1	MONSQDPSITSLPKGGGAI	TGMG	ALPTGPDG	MAALSPLPIS	AGRGVAPFTL	NN	60	
Qy	61	SGAGNSPFLG	WCNCN	MTIRRTTH	FGVPHYDETD	FTFLGPEGE	VLVADQPR	DESTLQ	IN 120
Db	61	SGAGNSPFLG	WCNCN	MTIRRTTH	FGVPHYDETD	FTFLGPEGE	VLVADQPR	DESTLQ	IN 120
Qy	121	LGATTTVTCY	RSLRSH	PSRLEYWQ	PKTKTKD	FTWLIYSPD	QVHLLGK	SPQARISN	PSQ 180
Db	121	LGATTTVTCY	RSLRSH	PSRLEYWQ	PKTKTKD	FTWLIYSPD	QVHLLGK	SPQARISN	PSQ 180

QY 181 TTOTAQWLLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLOAQRYLHIVYGNRTAS 240  
DB 181 TTOTAQWLLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLOAQRYLHIVYGNRTAS 240  
QY 241 ETLPGDGSAPSQADWLFYLVDFDGERSNLKTTPAFSTTGSWLCRODRFSRYEYGFIR 300  
DB 241 ETLPGDGSAPSQADWLFYLVDFDGERSNLKTTPAFSTTGSWLCRODRFSRYEYGFIR 300  
QY 301 TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLINLYVDESAIASTLVFVRVYRHEQDGNV 360  
DB 301 TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLINLYVDESAIASTLVFVRVYRHEQDGNV 360  
QY 361 VTLPPELAYODFSFRHHAWQPMDFANFNAIORQWLVDLKGSLGCLLYQDGAWWYR 420  
DB 361 VTLPPELAYODFSFRHHAWQPMDFANFNAIORQWLVDLKGSLGCLLYQDGAWWYR 420  
QY 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVINDGQGLDWVITGPGRLGYHSORPD 480  
DB 421 SAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVINDGQGLDWVITGPGRLGYHSORPD 480  
QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVVQ 540  
DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVVQ 540  
QY 541 SGDTITLVPVPGADPRKLVAFSDVLGSGQAHLYEVSAKTKVCPNLRGRFGQPIITLPGFSQ 600  
DB 541 SGDTITLVPVPGADPRKLVAFSDVLGSGQAHLYEVSAKTKVCPNLRGRFGQPIITLPGFSQ 600  
QY 601 PATEFNAQVYLAIDLDSGPTDLIVHTNRDLDFLNSKNGFAEPVTLRFPPEGLRFDHTC 660  
DB 601 PATEFNAQVYLAIDLDSGPTDLIVHTNRDLDFLNSKNGFAEPVTLRFPPEGLRFDHTC 660  
QY 661 QLQMAVQGLGVASLILSVPHMSPHHRCDLTNMKPMLNEMNNMNVHHTLRYRSSQF 720  
DB 661 QLQMAVQGLGVASLILSVPHMSPHHRCDLTNMKPMLNEMNNMNVHHTLRYRSSQF 720  
QY 721 WLDEKAAALTTGQFPVCVLPPIHTLWQTEDEISGNKLVTLRYARGAWGDERBFRG 780  
DB 721 WLDEKAAALTTGQFPVCVLPPIHTLWQTEDEISGNKLVTLRYARGAWGDERBFRG 780  
QY 781 FGYYEQTDSHOLAQGNAPERTPPALTKNWTATGLPVIDNALSTBYWRDDQAFAGFSRFT 840  
DB 781 FGYYEQTDSHOLAQGNAPERTPPALTKNWTATGLPVIDNALSTBYWRDDQAFAGFSRFT 840  
QY 841 TWQDNKDVLPEDDNGSRWYFNRLKGLLRLSELYGLDDSTNKVYPTVTTFERSQVRRLQ 900  
DB 841 TWQDNKDVLPEDDNGSRWYFNRLKGLLRLSELYGLDDSTNKVYPTVTTFERSQVRRLQ 900  
QY 901 HTDSRYPVLWSSVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPAI 960  
DB 901 HTDSRYPVLWSSVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPAI 960  
QY 961 NLYPDTLPDKLLANSYDDQQRLLTYQQSSWHLTNNTVRLGLPDSRSDIFTYGAEN 1020  
DB 961 NLYPDTLPDKLLANSYDDQQRLLTYQQSSWHLTNNTVRLGLPDSRSDIFTYGAEN 1020  
QY 1021 VPAGLNLLELLSDKXNSLIADKPRLYLQCKTAYTDGQNTTPTQTPRQALIAFTETTVF 1080  
DB 1021 VPAGLNLLELLSDKXNSLIADKPRLYLQCKTAYTDGQNTTPTQTPRQALIAFTETTVF 1080  
QY 1081 NQSTLSAPNGSIPBDKLSTTLEQAGYQCTNYLFPRTGEDKVVAAHGYTDYGTAAQFWRP 1140  
DB 1081 NQSTLSAPNGSIPBDKLSTTLEQAGYQCTNYLFPRTGEDKVVAAHGYTDYGTAAQFWRP 1140  
QY 1141 QKQSNLTQTKITLWADANYCVVQTRDAAGLITTSKAYDMRFLTPVQLTDINDNQHLITL 1200  
DB 1141 QKQSNLTQTKITLWADANYCVVQTRDAAGLITTSKAYDMRFLTPVQLTDINDNQHLITL 1200  
QY 1201 DALGRPITLRFWGTENGWQYSSPEKASFPSPDVNAAIIEIKPLPVAQCVQVAPESWM 1260  
DB 1201 DALGRPITLRFWGTENGWQYSSPEKASFPSPDVNAAIIEIKPLPVAQCVQVAPESWM 1260

QY 1261 PVLISQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLLNNGPRLP 1320  
DB 1261 PVLISQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLLNNGPRLP 1320  
QY 1321 PHSITLTTDTRDYHDPEQOIIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLINVOHT 1380  
DB 1321 PHSITLTTDTRDYHDPEQOIIRQOVVFSDFGRLLOAAARHEAGMARQORNEGSLINVOHT 1380  
QY 1381 ENRWAVTGRTEYDNKNGQPIRTYQPYFLNDMRYVNSDSARQSEKAYADTHVYDPIGREIKV 1440  
DB 1381 ENRWAVTGRTEYDNKNGQPIRTYQPYFLNDMRYVNSDSARQSEKAYADTHVYDPIGREIKV 1440  
QY 1441 ITAKGWFRRTLTFTPFWFTVNEDENTAAAEVKVKM 1474  
DB 1441 ITAKGWFRRTLTFTPFWFTVNEDENTAAAEVKVKM 1474  
RESULT 2  
Q7N7Y2 PHOLL PRELIMINARY; PRT; 1475 AA.  
AC Q7N7Y2 PHOLL PRELIMINARY; PRT; 1475 AA.  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Insecticidal toxin complex protein TcdB2.  
GN Name=tcdB2; OrderedLocusNames=plu0969;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
RA Duchaud E., Rusniok K., Frangeul L., Buchrieser C., Givaudan A.,  
RA Taourit S., Bosc S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,  
RA Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,  
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunat F.,  
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
luminescens".  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571862; CAE13264.1; -; Genomic\_DNA.  
DR Photolist; plu0969; -.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integrin complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR003284; Integrin alpha.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF03534; SpvB; 1.  
DR PRINTS; PR01341; SALSPVBPROT.  
DR Complete proteome.  
SQ SEQUENCE 1475 AA; 165779 MW; 24945174CDADAD22 CRC64;  
Query Match 92.2%; Score 7285.5; DB 2; Length 1475;  
Best Local Similarity 91.7%; Pred. No. 0;  
Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;  
QY 1 MNSQDFSITELSLPKGGGAIATGMGEALTPPTGPGMAALSPLPISAGRGYAPFTLNN 60  
DB 1 MNSQDFSITELSLPKGGGAIATGMGEALTPAGPGMAALSPLPISAGRGYAPSLALNN 60  
QY 61 SGAGNSPGLGWCNCVMITRRTHFGVPHYDETFTFLGPEGEVLVVAQPDDESTFLOGIN 120  
DB 61 SGAGNTFGLGWCNCVMITRRTHFGVPHYNETDTFLGPEGEVLVVAQPDDESTFLOGIN 120  
QY 121 LGATFTVTGYSRLESFHSRLLEYWQPKTTGKTDFWLIYSPDQVHLLGKSPQARISNFSQ 180  
DB 121 LGATFTVTGYSRLESFHSRLLEYWQPKTTGKTDFWLIYSPDQVHLLGKSPQARISNFSQ 180  
QY 181 TTOTAQWLLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLOAQRYLHIVYGNRTAS 240  
DB 181 TTOTAQWLLLEASVSSRGEIYYQYRAEDDTGCEADEITHHLOAQRYLHIVYGNRTAS 240

Db 181 TTQTAQWLLASVPHGEIYYQYRAEDNLDCTNETITLHPQATAQRYLHVHYGNRTAS 240  
Qy 241 ETLPLDGSAPSQADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEYGFIR 300  
Db 241 ETLPLNGRAPSQADWLFYLVDFYDYGERSNNLKMPPATATGWLRCRODRFSRYEYGFALR 300  
Qy 301 TRRLCROVLMYHHLOALDSKITEINGPTLVSRLLIYNDESIAISTLVFVRVHQDGNV 360  
Db 301 TRRLCROQILMYHCIALDNKIKENHGPTLVSRLLIYNDESIAISTLVFVRVHQDGTG 360  
Qy 361 VTLPLELAYODFSRPHAHQPMQDVLNANFNAIQWOLVDLKGRLGCLLYQDQKAWMYR 420  
Db 361 VTLPLELAYODFSRPHAHQPMQDVLNANFNAIQWOLVDLKGRLGCLLYQDQKAWMYR 420  
Qy 421 SAQRLBIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDGLDWVITGPGLRGYHSQRPD 480  
Db 421 SAQRLBIGSDVITWIKIQLSVIPSLQSNASLVINDINGDGLDWVITGPGLRGYHSQRPD 480  
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRODGFAGKDVQV 540  
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRODGFAGKDVQV 540  
Qy 541 SGDIITLVPAGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGOPITLPGFSQ 600  
Db 541 SGDIITLVPAGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGOPITLPGFSQ 600  
Qy 601 PATEFPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFABPVTLPPEGLRFDHTC 660  
Db 601 PEAIFPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFABPVTLPPEGLRFDHTC 660  
Qy 661 QLMADVQGLGVASLILSVPHMSPHHWRCDLTNMKPWLNMNNMNGVHHTLRYRSSQF 720  
Db 661 QLMADVQGLGVASLILSVPHMSPHHWRCDLTNMKPWLNMNNMNGVHHTLRYRSSQF 720  
Qy 721 WLDEKAAALTTGTPCVCLPPIITLMQTEDEISGNKLVTLRYARGADGGRERFRG 780  
Db 721 WLDEKAAALVAGTPCVCLPPIITLMQTEDEISGNKLVTLRYARGADGGRERFRG 780  
Qy 781 FGYYEQDTSQLOAGNAPERTPPALTNKNWYATGLFVIDNALSTYWR-DDQAFAGFSRPF 839  
Db 781 FGYYEQDTSQLOAGNASERTPPAMTKNKNWYATGLFVIDNTLSTBYWGDNDQAFAGFSRPF 840  
Qy 840 TTWQDNKDVPLTPEDDSRYWFRNALQGLRLSRLYGLDSTNKHVPYVTVTEFRSQVRL 899  
Db 841 TTWQDGKIDLLTPEDDSRYWFRNALQGLRLSRLYGLDSTNKHVPYVTVTEFRSQVRL 900  
Qy 900 QHTDSRYVPLWSSVVERSVHYRIASDPQSONITLSSDRFGQPLKOLSVOYPRROOPA 959  
Db 901 QQAQNYVPLWSSVVERSVHYRIASDPQSQDITLSSDILFQPLKQVSVQYPRRKOPA 960  
Qy 960 INLYPDTLPDKLANSYDDQORQLRLTYQSSWHLTNNTVRLGLPDPSTRSDIFTYGA 1019  
Db 961 ISPPYDPLPDELLANSYDEQORQLRLTYQSSWHLTDNTVRLGLPDPSTRSDIFTYGA 1020  
Qy 1020 NVPAGLNLLELLSKNSLIADKPVEYLGQOKTAYTDGQNTPTLQTPTRQALIAFTETTV 1079  
Db 1021 NVPAGLNLLELLSKNSLIADKPVEYLGQOKTAYTDGQNTPTLQTPTRQALIAFTETTV 1080  
Qy 1080 FNQSTLAFNGSIIPSKLSTLEAGYQOONYLFRPCEGKVVVAHHGYTDYGTAAQFWR 1139  
Db 1081 FSQSTLAFNGSIIPSKLSTLEAGYQOONYLFRPCEGKVVVAHHGYTDYGTAAQFWR 1140  
Qy 1140 PQKQNTQLTKITLIWDANYCVVVQTRDAAGLTTSAKYDWRFLTPVQLTDINDNQHLIT 1199  
Db 1141 PQKQNTQLTKITLIWDANYCVVVQTRDAAGLTTSAKYDWRFLTPVQLTDINDNQHLIM 1200  
Qy 1200 LDALGRPITLRFWGTENGKMTGYSSPEKASPSPSDNNAAIELKPLPVAQCQVYAPESW 1259  
Db 1201 LDALGRPITLRFWGTENGKMTGYSSPEKASPSPSDNNAAIELKPLPVAQCQVYAPESW 1260  
Qy 1260 MPVLISQKTFNLAQDQWKLNYRIITEDGRCITLAYERWVOSOKAIPOLISLANNPRL 1319  
Db 1261 MPVLISQKALNRLAEDQDKLNYRIITEDGRCITLAYERWVOSOKAIPOLISLANNPRL 1320

Qy 1320 PPHSLTLTTDRYDHPDPEQIIRQQVYVSDGFGRLLLQAAARHEAGMARQNRNEDGSLIINVQH 1379  
Db 1321 PPHSLTLTTDRYDHPDPEQIIRQQVYVSDGFGRLLLQAAARHEAGMARQNRNEDGSLIINVQH 1380  
Qy 1380 TENRWATVGTEDYDNKGPRTYQPYFLNDWRVYVSDSARQEKAYADTHVYDPIGREIK 1439  
Db 1381 TENRWATVGTEDYDNKGPRTYQPYFLNDWRVYVSDSARQEKAYADTHVYDPIGREIK 1440  
Qy 1440 VITAKGWRFTLTPFWFTVNEDENDTAAEVKKVQM 1474  
Db 1441 VITAKGWRFTLTPFWFTVNEDENDTAAEVNEVM 1475

## RESULT 3

Q7N720 PHOLL  
ID Q7N720 PHOLL PRELIMINARY; PRT; 1476 AA.  
AC Q7N720;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Insecticidal toxin complex protein TcdB1.  
GN Name=tcdB1; OrderedLocusNames=plu0961;  
OS Photorhabdus luminescens (subsp. laumondii).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=141679;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=TT01;  
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
TAourit S., Bours S., Bouraux-Eude C., Chandler M., Charles J.-F.,  
Dassa E., Deros R., Derzelle S., Freysinet G., Gaudriault S.,  
Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,  
Zouine M., Glaser P., Boenare N., Danchin A., Kunet F.,  
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
luminescens.",  
RL Nat. Biotechnol. 21:1307-1313(2003).  
DR EMBL; BX571862; CAB13256.1; -, Genomic\_DNA.  
DR Photolista; plu0961; -.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integrin complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR004113; Integrin alpha.  
DR InterPro; IPR003284; Sal\_SpV8.  
DR Pfam; PF01839; FG-GAP; 2.  
DR Pfam; PF03534; SpvB; 1.  
DR PRINTS; PR01341; SALSPVBPROT.  
KW Complete proteome.  
SQ SEQUENCE 1476 AA; 165737. MW; 06AF0EB945B5728 CRC64;

Query Match 76.8%; Score 6065.5; DB 2; Length 1476;  
Best Local Similarity 76.3%; Pred. No. 0;  
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

Qy 1 MNSQDSFISITELSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRYAPFTLNYN 60  
Db 1 MNSQTSFVAELSPLKGGGALTGMGEALTPTGPDGMAALSPLPISAGRYSPSLTLYNS 60  
Qy 61 SGAGNSPFLGWDQCNVMTIRRTHTFGVPHYDDETFILGPEGEVLVA-----DOPRES 114  
Db 61 SGAGNSPFLGWDQCNVMTIRRTHTFGVPHYDDETFILGPEGEVLVAALNENQADIRSES 120  
Qy 115 TLQGINLGAFTVTGYSRLSHESFSLRYEYQPKTTGKTDFWLIYSPDQVHLLGKSPAR 174  
Db 121 SLQGINLGEFTVTGYSRLSHESFSLRYEYQPKTTGKTDFWLIYSPDQVHLLGKSPAR 180  
Qy 175 ISNPSQTTQTAQWLLASVSPSRGEQIYYQYRAEDDTCCEADEITHHLOAQRVLIHYV 234  
Db 181 ISNPLNYSQTAQWLLASVSSHGQIYYQYRAEDTNCETDEFTAHNPATVQRYLOAVHY 240  
Qy 235 GNRSTASTLPGLDGSAQSADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYE 294

```
Db 241 GNLTA5VFPTLNGDDPLKSGWLECLVDFDGERKNSLSEIPFPFKASSLWLCRDRFSRYE 300
Qy 295 YGFEIRTRRLCRQVIMYHHLQALDSKITEHNGPTLVSRLLINLYNDESAIATSLVFRVRVGH 354
Db 301 YGFELRTRRLCRQILMFHRLQTLGSGQAGDDEPALVSRKLLDYDENAVISTLVSVRRIGH 360
Qy 355 EQDGNVVTLPLELAYQDFSPRHHAHQPMQDVLANFNAIORWQVLVDLKGESLPGLLYQDK 414
Db 361 EDNNTVTSLPLELAYQDFPEQKARQWQSMQDVLANFNAIORWQLLDLKGEGVPGVLYQDR 420
Qy 415 GAWMYRAQRLGEIGSDAVTWEKMOPLSVLPSLQNASLVDINDGQGLDQWVITGPGLRGY 474
Db 421 NGWYRQAQQAQGEEMWATWGRKQLLPITPALQDNASLMDINDGQGLDQWVITGPGLRGY 480
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMAGLSLVLIGPKSVRLYANTRDGFAG 534
Db 481 HSQHPDGSWTRFTPLDALPIEYSHPRQAQLADLMAGLSLVLIGPKSVRLYANNRDGFTQ 540
Qy 535 GKDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLYVEVATKVTCPNLRGRGFGQPIIT 594
Db 541 GRDVVQSGDITLPLFGADARKLVAFSVLGSGQAHLYVEVATQVTCWPNLIGHGRFGQPIIT 600
Qy 595 LPGFSQATBFPNPAQVYLAIDLQSGGPTDLIVYHTNRDLDFLNKSGNGFAEPVTLRPPEGL 654
Db 601 LPGFSQADNFPNPRVHLADLQSGPADLLIVYHTDRLEIFKNESGNSFAKFTLRFPDGL 660
Qy 714 RPDHTCQLQADVQGLGVASLILSVPHMSPHHWRCDLTNMPMLNEMNNMNGVHHHTLRY 714
Db 661 RFDDTCQLQADVQGLGAVSLILSVPHMAPHHWRCDLTNAKPMLLSEMMNMGAAHHTLHY 720
Qy 774 RSSQFWLDEKAAALTGQTPVCVLPPIHTLWQTETEDISGNKLVITLRYARGAWDR 774
Db 721 RSSQFWLDDKAAALATGQTPVCVLPFPVHTLWQTETEDISGNKLVITLRYARGAWDR 780
Qy 833 ERPRGRGYEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWR--DQAF 833
Db 781 ERPRGRGYEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWR--DQAF 840
Qy 893 GFSRPTTWQNDKDVPLTPEDDNRSYRWFNRALQGLRLSRLYGLDDSTNKHVPYVTVTEFR 893
Db 841 GFTPRFTWKEGKDVPAATPENNDDLYNFRNALQGLRLSRLYGLDDSEQQNIPYTVTESR 900
Qy 953 SQRRLQHTSRYPVLNWSVSESNHYHYRIASDPQSNITLSSDFRFGQPLKOLSVOYP 953
Db 901 PQVRLQDGGTASVPLWASVVENRSHYRIIGDPQCNQDITLSSDQFGQPLKQVSNQYP 960
Qy 1013 RRQOPALNLPDTPDKLLANSYDDQORQLRLTYQSSWHHLTNNTVRLVGLPDPSTRSDI 1013
Db 961 RNRQPTNFPDTPDLTFLASSYDDQQLLRLTYQSSWHHLTNNTVRLVGLPDPSTRSDA 1020
Qy 1073 FTYGAENVAGGLNLELLSDKNSLIADDKPREYLGQOKTAYTDGQNTTPIQTPTROALIA 1073
Db 1021 FTYDAKQVPDGLNLEALCAENSLIADDKPREYLNQORTFTYTDGKNQAPLEIPTRQALIA 1080
Qy 1133 FTETTVNQSTLAFNGISDLSLSTLEAGYQOTWYLPRTCTGDKVWAHGYTYDGT 1133
Db 1081 FTETAUTESLSLAFDGGITPDELPGLITQAGYQOEPYLPRTCTGDKVWAHGYTYDGT 1140
Qy 1193 AAQFWRPQKQSNQTLTKITLIMDANYCVVYQTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1193
Db 1141 EAQFWRPVQARNLTLLTKITLQWDTYCVITQTDAGLTTLVANYDWRFLTPVQLTDIND 1200
Qy 1253 NOHLITLALGRPTLRFWGTENGKMTGYSSPEKASFPSPDNNAALELKLPLVPAOCQV 1253
Db 1201 NVHLITLALGRPVTORFWGTENGKMTGYSSPEKASFPSPDINDTALATGLPLVPAQCLV 1260
Qy 1313 YAPESWMPVLQKTFNLAEDQWKLNARLITEDGRICTLAYRWQSQKAIPLQLSL 1313
Db 1261 YAPDSWMPVLSQETFNLTUEEQWKLNARLITEDGRICTLAYRWQSQKAIPLQLSL 1320
Qy 1373 NNGPRLPHPSITLTTDRYDHDPEQOIRQVQVFSDFGRLQLQAAARHAGMARQNRDGS 1373
```

```
Db 1321 TNSIGLPHHNLTLTPDRYDRDSGQIHQVAFSDGFGRLQLQASVVRHEAGAWQRNQDGL 1380
Qy 1374 IINVOHTENRWATVGRTEYDNKQPIRTYQPYFLNDWRVYVNSDSAROEKAYADTHYD 1433
Db 1381 VTKMEDTKTAWVATGRTEYDNKQPIRTYQPYFLNDWRVYVNSDSAR--KGAYADTHYD 1438
Qy 1434 IGRREIKVITAKGFRRTLTFTPWFTVNEDENDTAAEV 1469
Db 1439 IGRIRVITAKGWLROSQYFPFWFTVSEDENDTAAV 1474

RESULT 4
Q93EP6_PHOUJ
ID Q93EP6_PHOUJ PRELIMINARY; PRT; 1476 AA.
AC Q93EP6;
DT 01-DEC-2001 (Tremblurel. 19, Created)
DT 01-DEC-2001 (Tremblurel. 19, Last sequence update)
DT 01-MAR-2004 (Tremblurel. 26, Last annotation update)
DE TcdB1.
GN Name-tcdB1;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAL18487.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
SQ SEQUENCE 1476 AA; 165138 MW; 8E6AC3D109911995 CRC64;

Query Match 76.38; Score 6031.5; DB 2; Length 1476;
Best Local Similarity 75.98; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Qy 1 MNSQFSITELSLPKGGGAIITGMEALITPGDGMALSLPLPISAGRGYAPFTLYN 60
Db 1 MNSQFSITELSLPKGGGAIITGMEALITPGDGMALSLPLPISAGRGYAPFTLYN 60
Qy 61 SGAGNSPFLGWDGNVMTIRRTTHFGVPHVDETDFTFLGPEGEVLVVA-----DQPD 114
Db 61 SGTGNSPFLGWDGNVMTIRRTTHFGVPHVDETDFTFLGPEGEVLVVAALNEAGQADIR 120
Qy 115 TLOGINLGAFTVTGVRSLSHFSRLYEQWTKTDFWLIYSPDGQVHLLGKSPQAR 174
Db 115 TLOGINLGAFTVTGVRSLSHFSRLYEQWTKTDFWLIYSPDGQVHLLGKSPQAR 174
Qy 121 SLQGINLGAFTVTGVRSLSHFSRLYEQWTKTDFWLIYSPDGQVHLLGKSPQAR 180
Db 121 SLQGINLGAFTVTGVRSLSHFSRLYEQWTKTDFWLIYSPDGQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQTAQMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHQLATAQRYLHIVY 234
Db 175 ISNPSQTTQTAQMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHQLATAQRYLHIVY 234
Qy 181 ISNPLNVNQTAAQMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHQLATAQRYLHIVY 240
Db 181 ISNPLNVNQTAAQMLLEASVSSRGEIYYQYRAEDDTGCEADEITHHQLATAQRYLHIVY 240
Qy 235 GNTASATLPGDLGSGAPSDWLFYLVDFYGERNNKLTTPAPSTTSGSWLCRDRFSRYE 294
Db 235 GNTASATLPGDLGSGAPSDWLFYLVDFYGERNNKLTTPAPSTTSGSWLCRDRFSRYE 294
Qy 241 GNLTA5VFPTLNGDDPLKSGWLECLVDFDGERKNSLSEIPFPFKASSLWLCRDRFSRYE 300
Db 241 GNLTA5VFPTLNGDDPLKSGWLECLVDFDGERKNSLSEIPFPFKASSLWLCRDRFSRYE 300
```



```
QY 291 SRVEGPEIRTRRLCROVLVHYHQLALDSKITEHNGPTLVSRLLIYNDESAIASTLVFVR 350
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 300 SRVEGPEVRTRRLCQVIMFHRFTALMAGEAGTSDAPELVGRLLLDYDRNASVTILTAAR 359
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 351 RVGHEQGNVVTLPPELALAYQDFSPRHHAHQPMVLANFNAIQRWLVDLKGSGLPGLL 410
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 360 QLSHEPDGSPITLPPELAWQRFNLEKMPWQRFDALDNFNSQQRYQLVDLKGSGPGL 419
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 411 YQDKGAWYRSQRLGEISDAVTEWKMQLSVIPSLQSNASLVLDINGDGLDWITGPG 470
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 420 YQDRGAWYKAPQKQEDGSNAVYDKTAPLFTPLPSQDNASLMDINGDGLDWITASG 479
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 471 LRGHVSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLGPKSVRLYANTRD 530
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 480 IRGHVSQPDGKWHFTFIPINALPVEYFHPSPQFADLTGAGLSDLVLGPKSVRLYANQRN 539
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 531 GFARQKDVQSGDITLPPVFGADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNPLGRGRFG 590
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 540 GWRKQDVQPGTGITLPTVGTDAKLVAFSDMLGSGQQHLEIRANRVTCWPNLGHGRFG 599
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 591 QPITLPGSPATFENPAQVYVADLDGSGPTDLIYVHTNRDLIFLNSGNGFAEPVILRF 650
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 600 QPLTLPGFNQPFENPERFLADIDGSGTDDIYVQSGSLIYLNOSGNQFDAPLTLAL 659
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 651 PEGLRFOHTCOLQADVQGLGVASILISVPHMSPHWRCDLTNNKPMWLLNEMNNMGVHH 710
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 660 PEGVOFONTCOLQADVQGLGVASILISVPHMSPHWRCDLTNNKPMWLLNEMNNMGVHH 719
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 711 TLRVSSSQFWLDEKAALTTGQTPVCVLPFPIHTLMQTEDEISGNKLVTLTLRYARGA 770
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 720 TLHYRSSAQFWLDEKQLTKAGKSPACVLPFPMHLLMHTEIQDEISGNRLTSVSYSHGV 779
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 771 WGREREPGRGVQEQTDSHQAQGNAPERTPPALTKNWTATGULPVIDNALSTYWR-DD 829
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 780 WDKEREPGRGVQEQTDTTTFESHGTAPEQAAPSLISWFAFGTDDVDSQLATEYQADT 839
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 830 QAFAGFSPFTWQ--DNKDVPLPEDDNRVYFNRAKLGOLLSELYGLDSTNKKVPY 887
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 840 QAYSGFETRYVWDMANQTDQAFTP-NETQRNWLTRALKGOLLKTELYGLDGTOSQTVPY 898
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 888 TVTFRRSVRRLQHTDSRYPVLWSVBSRNYHYERIASDPQCSQNTLTSSDRFGQPLKQ 947
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 899 TVSESRYQVRSIPVEKETELSAWTAIENRSYHYERIIISDPQLSQSLRQLQHDIFGQSLOS 958
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 948 LSVQYPRQQAQINLYPDTLPKLLANSYDQOQLRLTYQSSWHHLTNNTVRLGLPD 1007
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 959 VDIAMPREKPAVNPYPTLPEALFDDSYDDQQQLRLVRQKNSWHHLTDGSEWRLGLPN 1018
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1008 STRSDIFTYGAEVNPAGGLNELLSDKNSLTADDKPREYLGQOQTAYTDGQNTTPTLOTP 1067
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1019 AQRDVTYTDKTIPTTEGISLEVLKADGLADEKAAYVLQOQTFYTAGQSEVILEKPT 1078
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1068 RQALIAFTETTVFNQSTLAFNGSIPSKLSTLLEAGYQOQNTLYFPRTGDKVYVAHGG 1127
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1079 LQALVAFQETAMDDTSLQAYEGVIEEQLNTLTQAGYQQVARLENFTGSSPVMVARQG 1138
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1128 YTDYGTAAQFWRPQKQSTQTLTKITLIWDANYCVVUTRDAAGLTTSKATDWFRLFPVQ 1187
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1139 YTDYRDATQFWRPQTRNSLLTGKTLTWTDHHCVIIQTDAAAGLTQARYDYRFLFPVQ 1198
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1188 LTDINDNOHLITLALGRPITLRFWGTENGKMTGYSSPEKASFSPSPDVNAAILKKPLP 1247
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1199 LTDINDNOHLITLALGRVTTSRFWGTBAGATGYSN---QPFTPPDSVDKALALTGVLP 1255
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1248 VAQCVYAPESWMPVLSOKTFENRL---AEQDWQKLYNARIITTEDGRICTLAYRWRVOSQK 1304
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1256 VAQCLVAVDWSMPLLSLSQSPESAEALWAQLRAAHMVTEDGKVCALGKRGIGQON 1315
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1305 AIPOLISLANNPRLPHSHLITLTDYDHPDQOIRQOVVPSDFGRLLOQAARHEAGMA 1364
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1316 LTIQILSLIASIPRLPHVLGITTDYDNDPQQOQHQQTQVFSDFGRLQSSVRHEPGDA 1375
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1365 RORNEGDSLIINVO-----HTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDA 1418
```

```
Db 1376 WQKEDGGLVDANGALVSAATDRWAVSGRTEYDDKGQPVRTYQPYFLNWRVYVSDSA 1435
QY 1419 ROEKEAVADTHVYDPIGREIKVITAKGWFRRTLTFTPWFTVNEDENDTAA 1467
Db | : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1436 RD--DLFADTHLYDPLGREYKVIITAKKYLRKQYTPWFIIVSESDENDTAS 1482

RESULT 6
O85153 PHOLU
ID O85153 PHOLU PRELIMINARY; PRT; 1485 AA.
AC O85153;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Insecticidal toxin complex protein tcaC (toxin complex protein).
GN Name=tcaC;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA Bhattia R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens.";
RL Science 280:2129-2132(1998).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF046867; AAC38625.1; -; Genomic DNA.
DR EMBL; AF346497; AAL18451.1; -; Genomic DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR PROSITE; PS00626; RCCL_2; UNKNOWN 1.
SQ SEQUENCE 1485 AA; 166164 MW; 74228E27A9921EBD CRC64;
```

```
Query Match 58.2%; Score 4595.5; DB 2; Length 1485;
Best Local Similarity 57.8%; Pred. No. 4.1e-284;
Matches 861; Conservative 224; Mismatches 375; Indels 29; Gaps 11;

QY 1 MNSQDSITSLSPKGGGAIITGMGEALTPTGPDMAALSPLPISAGRGYAPAFITLNN 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MQDSPEVSIITLSPKGGGAIINGMEALNNAAGPDMASLSPLPLSTGRGTAPGLSLIYS 60

QY 61 SGAGNSPFLGWCDCNVMTIRRTTRHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDE 113
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 NSAGNPFPGIGWQGVWSISRRTQHGIPQYQVNDTFLSPQGEVMNIALNDQGPDIRQDV 120

QY 114 STLOGINLGATFTVTGRSLRLESFRLYEQPKT--TGKTDFWLIYSPDQGVHLGKSP 171
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121 KTLQGVTLPISTVTVYQARQLDIFSKEYWQASQGEGRA-FWLISSPDQLHLGKTA 179
```



QY 172 QARISNPSTTQTAQWLLBASVSSRGEQIYYQYRAEDDTGCEADEITHHQAQRYLHI 231  
DB 180 QACLANPQNDQOIAQWLLTETTPAGEHVSYYQYRAEDEAHCDNETHAPNVTQAQRYLVQ 239  
QY 232 VYVGNRTASETLPLGLDGSAPSQADWLVYLVFDYDGRSNNLTKPPAFST--TGSMLCRQDRF 290  
DB 240 VYGNIKPQASLVFNADNAPAEEMLFHLVDFHGERDTSLHTVPTWDAGTAQMSVRPDPF 299  
QY 291 SRYEYGFETRRLCRQVLMYHHLQALDSKITHNGPTLVSRLLIYNDESASIAITLVFR 350  
DB 300 SRYEYGFETRRLCRQVLMYHHLQALDSKITHNGPTLVSRLLIYNDESASIAITLVFR 359  
QY 351 RVGHODGNVTLPLLEAYQFSPRHAAHQPMDFLANFNAIQWQLVDLKGEGPLGL 410  
DB 360 QLSHESDGSPTQPPLLEAWQFDELEKMTQWQFALDNFNSQORYQLVDLKGEGPLGL 419  
QY 411 YODKAWYRSQAQLGEIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDQDWDWITGPG 470  
DB 420 YODKAWYRSQAQLGEIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDQDWDWITGPG 479  
QY 471 LRGYHSORPDGTSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 530  
DB 480 IRGYHSQPDGKWTHTFPINALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRD 539  
QY 531 GPKAKDVVQSGDIITLVPDAPRKLVAUSDVILGSGQAHLEVSATKVTCPNLRGRPG 590  
DB 540 GWRKEDVPQSTGITLPTVTGTARKLVAFSDMLGSGQQHLEIKANRVTCWPNLGHGRFG 599  
QY 591 QPITLPGFSOPATFNPAQVVLADLDGSGPTDLYVHTNRLDIFLKSGNGFAEPTVLRP 650  
DB 600 QPITLPGFSOPATFNPAQVVLADLDGSGPTDLYVHTNRLDIFLKSGNGFAEPTVLRP 659  
QY 651 PEGLFHDHDTCOLQADVQGLVASLILSVPHMSPHHRCDLTNMKPMLNEMNNNGVHH 710  
DB 660 PEGVQFQDNTCOLQADVQGLVASLILSVPHMSPHHRCDLTNMKPMLNEMNNNGVHH 719  
QY 711 TLRYSSQFMDLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTTLRYARGA 770  
DB 720 TLRYSSQFMDLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTTLRYARGA 779  
QY 771 WGRERERFGVYVOTDSHQAQNAQNAQNAQNAQNAQNAQNAQNAQNAQNAQNAQNAQ 829  
DB 780 WGRERERFGVYVOTDSHQAQNAQNAQNAQNAQNAQNAQNAQNAQNAQNAQNAQ 839  
QY 830 QAFAGFSRFTWQ--DNKDVPLTPEDDNRYSYFNRKALGOLLRELYGLDDSTNKHVPY 887  
DB 840 QAYSGFETRYTWDHNTQDAFTP-NETQWNLTRALKGOLLRELYGLDDSTNKHVPY 898  
QY 888 TVTEFRSQRRLQHTDSRYFVLMSSVBSRNYHYERIASDPQCSQNTLSSDRFGQPLKQ 947  
DB 899 TVTESRYQVRSIPVNMKETELSAWVTAIENRSYHYERIIITDPQFSQIKLQHDIFQSLQS 958  
QY 948 LSVQYPRQPAINLYPDTLPDKLLANSVDOORCLRTYQSSWHHLTNVTVRVLGLPD 1007  
DB 959 VDIAMPREKPAVNPYPTLPETLFDSSYDQQLRLVRQNSWHHLTNVTVRVLGLPD 1018  
QY 1008 STRSDIFYGAENVPAGNLLELSDKNSLADDKPREYLGQOKTAYTDGQNTTLPLOPT 1067  
DB 1019 AQRDVTYVYDRSKIPTGEGISLEILLKDDGLADEKAAVYLGQOKTAYTDGQNTTLPLOPT 1078  
QY 1068 QOALIAFTETTVFNOSTLSAFNGSTPSDKLSTLEQAGYQQTNYLFPRTGDKVVAHNG 1127  
DB 1079 LQALVAFQETAMMDTSLQAVEGVIEEQELNLTALTAQYQVQVARLNTFRSSPVWAARQ 1138  
QY 1128 YTDVCTAQAQFWRPQOSNTLTGKTLIWDANYCVVQTRAGLTSYAKYDWRFLTPVQ 1187  
DB 1139 YTDVCTAQAQFWRPQOSNTLTGKTLIWDANYCVVQTRAGLTSYAKYDWRFLTPVQ 1198  
QY 1188 LTDINDNQHILTLDALGRPITLRFWGTENGKMTGYSSPEKASFSPSPDVNAIIEKKPLP 1247  
DB 1199 LTDINDNQHILTLDALGRPITLRFWGTENGKMTGYSSPEKASFSPSPDVNAIIEKKPLP 1255  
QY 1248 VAQCOVYAPESMMPVLSQKTFNRL---AEQDWQKLNYNARIITEDGRICTLAYRRWVOSQK 1304

DB 1256 VAQCLVYAVDSWMPSSLSQLSQSQQEAEALWAQRAAHMITEDGKVCALSGKRGTSQHN 1315  
QY 1305 AIPOLISILNNGPRLPHSLTLTDDYDHDPEQIIRQOVFSDFGFRLLQAAARHEAGWA 1364  
DB 1316 LTIQLISILLASIPRLPHVLTITTDYDSDFOQQHQVTSFSDGFRLLQSSARHESGDA 1375  
QY 1365 FORNEDGSLIINVQ-----HTENRWAVTGRTEYDNKQPIRTYQPYFLNDWRVYVNSDA 1418  
DB 1376 WORKEDGLVVDANGVLVSAPTDTRWAVSGRTEYDKQPVRTYQPYFLNDWRVYVNSDA 1435  
QY 1419 ROKEAYADTHVYDPIGREIKVITAKGWFRTLTLPWTFVNEDENDTAA 1467  
DB 1436 RD--DLFADTHLYDPLGREGYKVIITAKYLRKLYTPWTFVNEDENDTAS 1482

## RESULT 7

QXSP56\_YERFR PRELIMINARY; PRT; 1433 AA.  
ID Q6XP56\_YERFR PRELIMINARY; PRT; 1433 AA.  
AC Q6XP56;  
DT 05-JUL-2004 (TremBLrel. 27, Created)  
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TremBLrel. 27, Last annotation update)  
DE Tc1P2.  
OS Yersinia frederiksenii.  
OG Plasmid pMYF1.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=29484;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=49;  
RA Dodd S.J., O'Callaghan M., Ronson C.W.;  
RT "Yersinia enterocolitica contains plasmid-encoded genes that show  
RT similarity to the tc family of insecticidal toxins.";  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY220492; AAP57764.1; -; Genomic DNA.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0008305; C:integrin complex; IEA.  
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.  
DR InterPro; IPR000413; Integrin\_alpha.  
DR InterPro; IPR003284; Sal\_SpVB.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF01839; FG-GAP; 2.  
DR PRINTS; PR01341; SALSPVBPROT.  
DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 1.  
KW Plasmid.  
SQ SEQUENCE 1433 AA; 157723 MW; 15FEDEFAL5118A4B CRC64;

Query Match 50.4%; Score 3984.5; DB 2; Length 1433;  
Best Local Similarity 52.8%; Pred. No. 4.3e-245;  
Matches 786; Conservative 187; Mismatches 432; Indels 85; Gaps 15;  
QY 1 MONSDFSITELSLPKGGGAIATGMEALPTGPGMAALSPLPISAGRYAPFTLNYN 60  
DB 1 MONRQDMAITTPSLPSGGAVTGLKIDIAAGPDGAATLSIPLVSPGRTYPTGTLYN 60  
QY 61 SGAGNSPGLGWCNVMITRRTHFGVPHYDEDTFLGPEGEVLVA-----DQPRDES 114  
DB 61 SRAGNGTFFGWSLGGGPAVQRTRSGAPAYDEDTFTGPDGGEVLVPAALTAAGTPTQAT 120  
QY 115 TLQGINLGATFTVGYRSRLESFESRLEYWQTKTKTDFWLIYSPGQVHLLGKSPAR 174  
DB 121 TLLGISFGGNSFVQVYRSRTEGSLRSLRWVPDAEATDFWVLYTPDGOVALLGNAGAR 180  
QY 175 ISNPSQTTQTAQWLLBASVSSRGEQIYYQYRAEDDTGCEADEITHHQAQRYLHIYV 234  
DB 181 ISNPARPTQAVLWLMSSVSLTGBQYVYQYRAEDDTGCEADEITHHQAQRYLVAWY 240  
QY 235 GNRPTASETLPLGLDGSAPSQADWLVYLVFDYDGRSNNLTKPPAFSTTGS--WLCRQDRFSR 292  
DB 241 GNRQARALPAL-MSTFSDMSWLFVQVFDYGERSTALSEPAPWQTPGSGEWRRCQDCFS 299



Qy	61	SGAGNSPFLGWD	CNVMTIRRRTHFGVPHYDETD	TFGLPGGEVLWA	-----DQPRDES	114			
		:	:	:	:				
		:	:	:	:				
Db	61	SRSNGNPGF	IGWGIGGA	AVQRRTRNGA	PTDYDDTDEFTGPDGEVL	PAALTAAGTQEARQAT	120		
		:	:	:	:	:			
		:	:	:	:	:			
Qy	115	TLOGINLGAT	PTWYGBSR	LESFESR	LEYQPKTTGKTDFWL	LYSPDGOVHLLGKSPQAR	174		
		:	:	:	:	:			
		:	:	:	:	:			
Db	121	SLUGINPGGS	FNVQYRSR	TEGSLUSRL	WMLPADETEFWLV	YTPDGOVALLGRNAQ	180		
		:	:	:	:	:			
		:	:	:	:	:			
Qy	175	ISNPSOTTQ	AOMLLEAS	VSRSRGEIQY	QYRAEDDTC	EADETHHLQATAQRYLHIVVY	234		
		:	:	:	:	:			
		:	:	:	:	:			
Db	181	ISNPTA	PTQAVWL	MESSVSLTG	OMYQYRAEDD	DCDEARDAHPQAGNQRYPAVWY	240		
		:	:	:	:	:			
		:	:	:	:	:			
Qy	235	GNTASTL	PLGLDGS	AFSQADWL	PVLPDYGERS	NNLKTTPAFSTTGS--WLCRQDRFSR	292		
		:	:	:	:	:			
		:	:	:	:	:			
Db	241	GNRQAART	LPAAL-VST	PSMDSWLF	ILVPDYGERS	SVLSEAPAWQTPGSGEWL	CRQDCFSG	299	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	293	YEYGFETR	RRLCRQV	LMYHHLQ	ALDSKITEHNG	PTPLSVRLILNYDESA	TASTILVFRVY	352	
		:	:	:	:	:			
		:	:	:	:	:			
Db	300	YEFGNL	RRLCRQV	LMFYHLY	GLAGSGAND	PALISRLLLDYRES	PSLSLENHVQV	359	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	353	GHEODGNV	WTLPLPE	LAYQDPS	PRHHAHW	PMDVLANFN	AIQRMQLVDLKGEGPLGLLYQ	412	
		:	:	:	:	:			
		:	:	:	:	:			
Db	360	AYESD	GTSCAL	PALALGW	QTFPTPL	SAWQTRDD	MGKLSLLOPQYQVLDLNGEGVVGILLYQ	419	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	413	DKGAWY	TRSQR	LGEIGSD	AVTEKMO	PLSVIPSLQSN	ASLVDINDINGDGLDWTITGFLGR	472	
		:	:	:	:	:			
		:	:	:	:	:			
Db	420	DSGAWY	REPVR	QSGDDP	DPAVTW	GAAALPTMP	ALHNSGILLADLNGDGRLEWVVTAPGVA	479	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	473	GYHSOR	PDGSGW	TRTFPL	NALPVEYTH	PRAPQALD	MGAGLSDLVLIGPKSVRLVANTRDGF	532	
		:	:	:	:	:			
		:	:	:	:	:			
Db	480	GMYDRT	PGRDWL	HTPTLS	ALPVEY	APKAVLAD	ILGAGLTDMLIGPRSVRLYSGKNDGW	539	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	533	AKGQDV	VQSGDI	TLPVPG	ADPRKL	VAFSDVL	SGQAHLVEVSATKVTWNLGRGRGQP	592	
		:	:	:	:	:			
		:	:	:	:	:			
Db	540	NKGETV	QOQTER	LTLVP	EGVDPT	LVA	SDWAGSQQHLTEVRANGVRYWNLGHRGQGP	599	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	593	ITLPGF	SQAPATE	FNPAQ	YVLADL	DGSGFT	DLLIYVHTNRDLIDFLMKSGNGFAEPVTLRFFE	652	
		:	:	:	:	:			
		:	:	:	:	:			
Db	600	VNIPGFS	QSVTTF	NPDDQ	ILLADT	DGSGT	TDLIYAMS	DLVIFYNQSGNYFAEPHTLLLPK	659
		:	:	:	:	:			
		:	:	:	:	:			
Qy	653	GLAPD	HTCLO	QADVQ	GLGV	ASLILSVPH	MSPHHRCDLTNMPKFWLLNENNNMGVHHTL	712	
		:	:	:	:	:			
		:	:	:	:	:			
Db	660	GVRDY	RTCSL	QVADIQ	GLGV	PSLLTVP	HVAPHWVCHLSADIPKFWLLNMMNMGARHAL	719	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	713	RYSSS	OFWLD	EKAAAL	TTGOT	PVCYLP	PPHITLWQTEDEISGNKLVTLLRYARGAWD	772	
		:	:	:	:	:			
		:	:	:	:	:			
Db	720	HYRSS	VOF	WLD	EKAAAL	AGSSPAC	YLPFTLTLWRSVQODEITGNRLVSDVILYRHGVMD	779	
		:	:	:	:	:			
		:	:	:	:	:			
Qy	773	GREREP	FGFV	GEOTD	SHOLA-QON	APERTPP	ALTKWYATGLPVIDNALSTBYWRDD-Q	830	
		:	:	:	:	:			
		:	:	:	:	:			
Db	780	QGEREP	FGFV	ELRDT	DLASQ	GTATLS	MPSVSRNWIATGVPADVERLPEIYWQNDAA	839	
		:	:	:	:	:</			

Qy	1127	GYTDYGTAAQFWPQKQSNLTGKTLIWDANYCVVQTRDAAGLTTSKAYDWRFITPV	1188
Db	1134	GYTYAGAEHFWLPLFLRDSMLTGPVTVTRDAYDCVITQWQDAAGIVTTADYDWRFITPV	1193
Qy	1187	QLTDINDNQHLLITLDALGRPITLRFWGTENGKMTGYSSPEKASFSPSDVNAAILKPKL	1246
Db	1194	RVTDPNDNLQSVITLDALGRVITLRFWGTENGATGYSD---ATLSVPDGAAGAAALATAPL	1250
Qy	1247	PVAQCQVYAPESPMVPLSQKTFNRLAEQDWQKLYNARIITEDGRICLTAYRRVWQSKAI	1306
Db	1251	PVAQCLVYVTDWS-----GDNDNEK-----	1270
Qy	1307	PQLISLLNNGPRLLPHSLTLTTDRYDHPDQQLRQVVFSDGFGRLLLQAAARHEAGWARQ	1366
Db	1271	-----MPPHHVVLAATRYDSDTGCGVRRQVTTSDGFGREQLQSTRQAEGNWQ	1318
Qy	1367	RNEDGSLI-----INVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRVYVNSDSAR	1419
Db	1319	RGRDGLKLVATSDGLPVTVA--TNFRMAVTGRAEYDNKGLPVRVYQPYFLDSWQVVSDDSR	1377
Qy	1420	QEKNAVADTHVYDPIGREIKVITAKGWFRTLETFWFTVNEDENDT	1465
Db	1378	Q--DLYADTHFYDPTAREWQVITAKGERRQVLYTPWVVSSEDENDT	1421

RESULT 9

Q6QDZ4\_YERPS PRELIMINARY; PRT; 1489 AA.

AC Q6QDZ4\_YERPS Q665G5;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Y0185-like protein (Insecticidal toxin complex)

GN Name: tca; Ordered locus names: YPB3553; ORF names: YPP3678;

OS *Yersinia pseudotuberculosis*.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

NCBI\_TaxID=633;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=IP32953;

RA Bernardes Pinheiro V.B., Ellar D.J.;

RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.

[2]

RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=IP32953 / Serotype I;

RK PubMed=1538859; DOI=10.1073/pnas.0404012101;

RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,

RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,

RA Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,

RA Simonet M., Chenail-Francisque V., Souza B., Dacheux D., Elliott J.M.,

RA Derbise A., Hauser L.J., Garcia E.

RT "Insights into the evolution of *Yersinia pestis* through whole-genome

RT comparison with *Yersinia pseudotuberculosis*.";

RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).

DR EMBL; AV545599; AAS66065.1; -; Genomic DNA.

DR EMBL; BX936398; CAH22791.1; -; Genomic DNA.

DR GO; GO:0005737; C.cytoplasm; IEA.

DR GO; GO:0008305; C.integrin complex; IEA.

DR GO; GO:0007160; P.cell-matrix adhesion; IEA.

DR InterPro; IPR000413; Integrin\_alpha.

DR InterPro; IPR000408; Reg\_chr Condens.

DR InterPro; IPR003284; Sal\_SpVb.

DR Pfam; PF01839; FG-GAP; 1.

DR Pfam; PF03534; SpvB; 1.

DR PRINTS; PR01341; SALSVPBPROT.

DR PROSITE; PS00626; RCC1\_2; UNKNOWN\_1.

DR Complete proteome.

SQ SEQUENCE 1489 AA, 167592 MW, CEA5B7C0A04B7B44 CRC64;

Query Match 46.0%; Score 3638; DB 2; Length 1489;  
Best Local Similarity 50.1%; Pred. No. 6.3e-223;  
Matches 756; Conservative 214; Mismatches 476; Indels 62

```
QY 1 MONS-QDFSTIELSLPKGGGAIITGMEALTTGPDGMAALSPLPIAGRGYAPAFITLY 59
DB :|||:::VA:VAPLPLPKGGGAIITGMSLGPISGMAITLPLPIAGRGYAPSLTLY 60
QY 60 MSGAGNSPFLGWCNVMYTIERRTHFGVPHYDETDITFLGPEGEVLVADQPD--ESTLQ 117
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 120
QY 61 SSGGNGPFLGWLGTWAIIRRTNAQVPRIDEVDFLAPNGEVWVVAADQGNIERTEQ 120
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 120
QY 118 GINLGATFTVTGYSRSLRSHESFRLEYWPKT-TGKTDPLWLYSPDGQVHLGLKSPQARIS 176
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 179
QY 121 SLN-GEQFSVIRLPRIEGNEFHRIEYRPRTNNSQAPFLVHSSDGQKHGLGSASARIA 179
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 179
QY 177 NPSQTTQATQWLLSEASVSRGEQIYYQRAEDDTGCEADEITHLQATQRYLHVYGN 236
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 239
QY 237 RTASETLPLGLDGSAPSQADMLFYLVDYGERSNLKTTPAPSTTGSWLCRODRSRYEYG 296
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 298
QY 240 REVAHELKLT-QRPAPTSWFLPSLFDHGEVSNIAEQVPVILKGSWNFRQDAFRPSCG 298
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 298
QY 297 FEIRTRLCROVLMYHLQALDSDKITEHNGP----TLVSRILNYDSATIASTLTVFVRV 352
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 353
QY 353 GHEQDGNVVTLPPLLEYADQSPRHHAHQPMQVLANFNATQRMQLVDLKEGGLPGLLYQ 412
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 411
QY 354 AHEPDGTRKSLPLPLEFDYQDFSTDALGWLQPLTDWAEFN--YQYQMDVLNKGMPGMLYQ 411
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 411
QY 413 DKGAWYRSARQLGEIG-SDAVTWKMQPLGVIPLSQSNASLVINDGQGLDWITGPGL 471
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 469
QY 412 DSGHWIYRPPVR--QPGTADGITGAAQQLPSLPAMRENAMLDINGDGLDWISQPL 469
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 469
QY 472 RGYH-----SQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLY 525
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 529
QY 470 AGYFSRDPDLSDRDPDLSWTQFPLSTLPAEYFHPQALVDLAGSLGSLALIGPKSVRVY 529
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 529
QY 526 ANTBDGPAKXDVQSGDITLIPVCGADPRKLVASDVLGSGOAHVSVATKTCWNLG 585
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 589
QY 530 TNLGDSFAAATQVQVDDITLPLPGAHEFTLVAFSDVMGSGQHLVIRHNSVTCWNLG 589
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 589
QY 586 RGRFQPIITLPGFSPATFENPAOVLADLDSGSGPTDLIYVHTNRDLIFLNKSGNGFAEP 645
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 649
QY 590 HGRGHPLSLGFGNQHIQFNFLAYLADIUGSGTIDLIIYATASQLLIYRQSGNRFAEP 649
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 649
QY 646 VTLRPFEGRLDHTCQLQMAVQGLGVASLILSVPHMSPPHWRCDLTNMKFWLLNEMNN 705
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 709
QY 650 VAIALPTGIRFDNSQLSLADIQGLGVASIMLSVPHPTTQHWRYDFVASKPEYLLCTNN 709
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 709
QY 706 MGVHTLRYRSSQFWLDEKAAALTGTGTPCYLPPPIHTLWQTEDEISGNKLVTLR 765
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 769
QY 710 MGASQLFYRSAQFWLDEKAAQKQKSLACQLFPPIHLLAQTTQLDDEITGNSLSQTAR 769
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 769
QY 766 YARGAWDGRERFGRGVVEQTDHQALQAQGNAPERTPPALTKNMYATGLPIDNALSTEY 825
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 829
QY 770 YFHGFYDGVQREFCGFGRVDITLDTTSAQGAERTATPKSRFHTGRADNETLWQSEY 829
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 829
QY 826 WR-DDQAPAGSPRFTTWQDNK--DVPLTTPEDDSRYWFNALGQLRLSRLYGLDDSTN 882
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 889
QY 830 WQGDQDQYPLPLTRLTKEINDTQGGDLSLSELDNDQTFWLHRALGSLRLSRLYGLDDSKL 889
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 889
QY 883 KHVPYTVTFRSOYRRLQHTDS--RYPVLMSVSVESRNYHYERTASDPQCSQNTLSSDR 940
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 949
QY 890 ATQPYSVNSRYQVRQIQSSADGISSPVALPMVLEQUSYHYERIAQDPQCSQQLVHCNE 949
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 949
QY 941 FGQPLKOLSVQPRRQPAINLYPDTLPDKLLANSYDQQRQLRLTYQQSSWHHLTNVT 1000
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1008
QY 950 FGHPLHSVTINYPRDKARIISY-SWLAKHWDNSFDEQQQLRLITESQQSYVHSIDKF 1008
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1008
QY 1001 RVLGLPDSSTRSDIFTYGAEVNPAGGLNLELSDKNSLITADDPKREYLQOQKTAITDQNT 1060
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1066
QY 1009 YVLGLPAGQRSDVLTYPNFPVFTAGIHWEELQQPEGLLGTAKERTFAQQQQQVFT--SDT 1066
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1066
```

```
QY 1061 TPLQTPRQALIAFTETTTFVNSTLSAFNGSIPSDKSLSTLLEQAGYQOYNTYLPFRGTGDK 1120
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1119
QY 1121 VWAHGHYTDYGTAAQAFWRPQKQSNLTQTKITLIWDANYCVVVQTRDAAGLTTSAKYDW 1180
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1179
QY 1181 RLTPVOLTINDNQHLITLDALGRPITLRFWTE---NGKM--TGYSSEKASFSPPS 1234
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1239
QY 1235 DVNAAILKXP-LPVAQCVYAPSWMPVLSQ---KTFNRLAEQDWOKLYNARIITEDGR 1290
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1299
QY 1291 ICTIAYRWV-QSOKAIPQLISLNNGP-RLPPHSLTLTTDRYDHDDE-QOIROQVVFSD 1347
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1359
QY 1348 GFGBELLQAAARHEAGMARQORNEGSLIINVQ-----HTENRWAVTGRTEYDNKGPIRT 1401
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1419
QY 1402 YQPYFLNDWRVSNDSARQEKAYADTHVDPDIGREIKVITAKGWFRTTLPFWFTVNE 1461
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1477
QY 1462 ENDTAAEV 1469
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1485
QY 1478 ENDTASEI 1485
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 1485

RESULT 10
Q693A4 YEREN PRELIMINARY; PRT; 1481 AA.
ID Q693A4 YEREN PRELIMINARY; PRT; 1481 AA.
AC Q693A4;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Putative insecticidal toxin complex protein.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=T83;
RA Tennant S.M., Skinner N.A., Joe A., Robins-Browne R.M.;
RT "Versinia enterocolitica biotype 1A insecticidal toxin complex
genes.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY647257; AAT90757.1; -; Genomic DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007150; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR000408; Reg_chir_condens.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR PROSITE; PS00626; RCCL_2; UNKNOWN 1.
SQ SEQUENCE 1481 AA; 136232 MW; 405D1D3D94868613 CRC64;

Query Match 45.4%; Score 3591; DB 2; Length 1481;
Best Local Similarity 49.5%; Pred. No. 6.3e-220;
Matches 738; Conservative 209; Mismatches 501; Indels 44; Gaps 22;

QY 3 NSQDFSTIELSLPKGGGAIITGMEALTTGPDGMAALSPLPIAGRGYAPAFITLYNSG 62
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 63
QY 4 NKQGVGAPLSPKGGGAIITGMEALTTGPDGMAALSPLPIAGRGYAPAFITLYNSG 63
DB :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||::: 63
```

Qy	63	AGNSPPFGLGWCDCNYMTIRRTHF	GPVHYDETDTFLGPEGEVLV	VADQPRDESTLQGINLG	122	
Db	64	SGNGPFGLGWNIGVLSHRSVTG	KVPVYDHRDQFIGTGV	LVAVSE-SNEISQDKRYAS	122	
Qy	123	ATFTVTGYSRLESHPSRLEWOP	KTTGK--TDFWLIYS	PDGQVHLLGKSPQARISNPSQ	180	
Db	123	ABFTITTYQRIEGGFDRIEFPWQ	KNTAAADLTSFWLIYSADGQ	QOHCGLGKSAQAHADPEH	182	
Qy	181	TTQTAQWLLBASVSRSRGEQIY	QVRAEDDTGCEADEITHLQ	ATAQRYLHVIVYGNTAS	240	
Db	183	PDHTAEWLLEESVSAMGEHCYD	QSEENDCNIDTAEGTAHPK	VCAQRYLQOVRYGNLTSS	242	
Qy	241	ETLPGLDGSAPSQADWLFYLV	FDYDGERSNNLKTTPAFST	TGSMWLCRODRFSRVEYGFETR	300	
Db	243	EVLVTYVKPSAKDKWLF	FTLVFDYGERTNTLSDAP	LQKPNGNWPCRDPSRVEYGFVR	302	
Qy	301	TRRLCROVLMYHHIQA	LDSDSKITEHNGPTLVSR	LILNYDESAIASTLVFVRV	GHEQDGNV 360	
Db	303	TRRLCQOVLMFHNHQLL	--SDDEKANDAAALTARU	LINYLISPYATQLTSSCQO	VAHEKDGTT 361	
Qy	361	VTLPELELAYODFSRHHAWQ	PMDLANFNAIQRMQLVDL	KGBGLPGLLYQDKGAWWYR	420	
Db	362	LTLPLPEFDQELTAETEAN	QWAPKSPSPY--YQV	VDLNGEGLPMLYQDSGSWHYH	419	
Qy	421	SAQRLGIGS--DAVTWEK	WOPLSVIPLOSNASI	VDINGDQLDWMVTG	PGLRGYHSQRP 479	
Db	420	PPVRTP--GSDDGVS	YQNAQILPNIPTL	RSASLMDINGDRLD	WVLSQPLGTGYYSRNP 477	
Qy	480	DGSWTRTPLNALPVE	THPPRAQADLMAGL	SDLVLGPKSVRLYANT	RDGFAKGDVV 539	
Db	478	DESWTQFTPLSALPAEY	FHPQAQIADLIGDGL	SDLALIGNSVRLYAN	QDGFAPAAKQVY 537	
Qy	540	QSGDITLVPVQADPRK	LVAFSDVLGSGOAH	IVEYSATKVTCPNL	KRGREGQPTTLPGFS 599	
Db	538	QDPAIALUPIGAN	SELVAFSDPLGSGOQ	HLMRVCHNSVTCW	NLNGTGTGTPVTLGPFN 597	
Qy	600	QPATEFNPAQVY	LADLDSGSGPTDILIY	VHTNRLDITFLNKS	NGFAEPVTLTFPPEGLRFDHT 659	
Db	598	QPIDSFPQHIWLAD	IDSGTGADLIYAKR	DHLAIYRQSGNAFAE	PLKLPLPHGVHYDNT 657	
Qy	660	COLQMAVQVGLG	VASLILSVPHMSPHWR	CDLTNMKPLLNNMNN	MNGVHHTLRYRSSSQ 719	
Db	658	COLTLADVQGLG	ITSLILSVPHIATRH	WRVYDFTTTPKPYLL	CTINNNGAENRLSYRSSVQ 717	
Qy	720	FWLDEKAAALT	TGQTPCYLPPFHTL	WQIETEDEISGNK	LVTTLRYARGAWDGRREER 779	
Db	718	FWLDEKEKAKEGR	RIACEMSFPMHLLART	QPDERTHNTLSQ	QAYIAHGFYDGTBRBFR 777	
Qy	780	GFGVVEQTD	SHOLAQGNAPERT	PALTKNWTATGLP	VIDNALSTEYWR--DDQAFAGFSPR 838	
Db	778	GFGRVDTL	DADARGAQEHPT	PLSCSWFHTGRAD	DETTRYQTEYWGQDDKAYQLGATR 837	
Qy	839	FTTW--QNKDV	PLVTPBEDDSRYW	FNFRALKQLLRSEL	YGLDDSTNKHVPYVTVEFRSQV 896	
Db	838	LTSFNTQTEAD	DLLEVTSSEQCYW	LYRALKGSLRLS	ELYGLDDIGLNEVPTINSARNQV 897	
Qy	897	RRLQHTDSR--Y	PVJWSVSVESRNYH	TERIASDQCQONIT	LSDDRFGQPLKQLSVQYPR 954	
Db	898	RQVRAADNTR	VSPTWMPISIEQL	SYHERIPQDPQC	SQSIVMRCNOYGHFLHTVAINYPR 957	
Qy	955	RQQPAINLY	PDTPDKLLANSY	DDQOQLRLTYQCS	SWHLLTNNTVRLVGLP	STDRSIDF 1014
Db	958	RPQOQVTPY	SWIATDH--WSSSY	DEBQRLRLRIS	EQSQSYQHVIOQDGLLLGLP	WQQRSDMQ 1016
Qy	1015	TYGAENVPAG	GLNLELLSDKNS	IADDKPREYLGQ	QOKTAYTDGQNTT	PLQTPTRQALIAF 1074
Db	1017	IYSADAVP	INGLNEYIVTKE	SLPTFIESYSQ	SEVFYHPEH-----VLQGLVIA 1068	
Qy	1075	TETTVFNQ	STLASFNGIS	PSDK-LSTTLE	QAGYQOOTYLPF--RTGBDKWVA	HGHYTDY 1131
Db	1069	SQSAEF				

Qy	1132	GTTAAQFWRPQKQNTQLTGKIITLWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVLQITDI	1139
Db	1129	HDASKFTYRAAQSRSLVAGATNLSDWSCTCAVTCITLADGSGITTTAQYDYRFTAPYLLTDI	1188
Qy	1192	NDNQHLITLTDALGRPITLRFWGTEN--GKMTGYSSPEKASFPSPSDVNAAI--ELKMKLPVPA	1249
Db	1189	NDNQHSLEFDALGRVISTRICGTEQIGNQVVGCGFTNPTPIADSDTDAIAKENEVIPVA	1248
Qy	1250	QCQVYAPESWMPVLQSQTFNRLAEQDW--OKLYNARIITEDGRICITLAYRRWV--QSQKAI	1306
Db	1249	QFFIYQALTWMPKLSPOLINDLFNQTFQDNLIQNHWMTEDGYLYALGLRRLWRKQNKSSI	1308
Qy	1307	PQLI-SLLNNGPRPLPHSLTLTTDRYDHD--PEQIRQOVVFSDFGFGELLOQAAARHEAGMA	1364
Db	1309	THAIESQLNNEQRHPHAMTVATDSYDRDAKEQHQQLIYVNDGLGRALQSAQRVDAGEA	1368
Qy	1365	RQRNEDGSL-----IINVOHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRWYVSNDS	1417
Db	1369	YIRQVNGQLQTGSGPVLNI--VKQRWASIGRVEYDNKGLVVRAYQPYFLDDWRWYISDDSDS	1426
Qy	1418	AROEKEAYADTHVYDPFIREIKVITAKGWRFRRTLTFTFWFTVNEDENDTAAEV	1469
Db	1427	AR--KOTYADTHIYDPLGREIKVITAKGYLRRLTQHFPFWFVISEDNDTAAEV	1476
RESULT 11			
Q8ZAV4 YERPE			
ID	Q8ZAV4 YERPE PRELIMINARY;	PRT; 1496 AA.	
AC	O8ZAV4: 074PN2;		
DT	01-MAR-2002 (TrEMBLrel. 20, Created)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)		
DT	01-FEB-2005 (TrEMBLrel. 29, Last annotation update)		
DE	Insecticidal toxin complex.		
GN	Name=tcaC; Synonyms=tcaC1; OrderedLocusNames=YP3866, YPO3678;		
OS	Yersinia pestis.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Yersinia.		
OX	NCBI_TaxID=632;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=CO-92 / Biovar Orientalis;		
RC	MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;		
RA	Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Hoiden M.T.G.,		
RA	Pringle M.B., Sebailha M., James K.D., Churcher C.M., Mungall K.L.,		
RA	Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.-M.,		
RA	Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,		
RA	Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,		
RA	Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,		
RA	Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.,		
RT	"Genome sequence of Yersinia pestis, the causative agent of plague."		
RL	Nature 413:523-527(2001).		
RN	[2]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=91001;		
RC	PubMed=15369893;		
RA	Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,		
RA	Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,		
RA	Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,		
RA	Yang H., Wang J., Huang P., Yang R.;		
RT	"Complete genome sequence of Yersinia pestis strain 91001, an isolate		
RT	avirulent to humans;"		
RL	DNA Res. 11:179-197(2004).		
DR	EMBL; AJ414158; CAC93148.1; -; Genomic DNA.		
DR	EMBL; AE017142; AAS64011.1; -; Genomic DNA.		
DR	PIR; AH0447; AH0447.		
DR	GO; GO:0005737; C:cytoplasm; IEA.		
DR	GO; GO:0008305; C:integrin complex; IEA.		
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.		
DR	InterPro; IPR000413; Interpro alpha.		
DR	InterPro; IPR000408; Reg_chx_Condens.		
DR	InterPro; IPR003284; Sal_SpVb.		
DR	Pfam; PF01839; FG-GAP; 1.		
DR	Pfam; PF03534; SpvB; 1.		

Qy	940	RFGOPKQLSVQYPRROCPAINLYPDTPLPKLLANSVDDQOROLRLTYQQSSWHHLTNWT	999
Db	954	EYGHPLSHVITNPPRRDKARISPY-SWLAKHWDHSHFDEQOQQOQRITTESQOOSYHHEISDK	1011
Qy	1000	VRVGLPDSRSDIFTFYGAENVPAAGLNLELLSDKNSLIADDPREYLGQCKTAYTDGON	1055
Db	1013	FYVGLGLPAGQSRSDVLTYPDNFVPTAGTHWEELQOPEGLLGTKAERTFTTAQOQVFT--SD	1070
Qy	1060	TTPLQTPTRQALIAFTTTFVFNQSTLSAFNGSIPSDKLSLTLEQAGYQQTNYLFPRTGED	1111
Db	1071	TIP-----GLVAYSQAEFDDQTLVALDELLPANERKQQLIKAGYQIAPRLPARTGET	1122
Qy	1120	KVWVAHHGYTDYTAAGFWRPQKOSNTOLGKTLIWDANYCVVQTRDAAGLTTSAKYD	1177
Db	1124	DIWVAQSGFTDYGDAERFYRPIQRSTQLVGKLTILEWDATCCAVSDIILADYSITHAEYD	1188
Qy	1180	WRFLTPVOLTDINDNOHLITLDALGRPTITLRFWCTE-----NGKM--TCYSSPERKASFSPP	1233
Db	1184	YRFITPYLLDINDNOHYIELDALGRVTSSRFAGTEIDPQNKVIETGFPSSIAEQFSPAP	1244
Qy	1234	SDVNAALTELKPK-LPVAQCQVYAPESWMPVLVSQKTFNRLAQQDW-----QKLYNARI	1288
Db	1244	NSVDKALSLENTRIPVAQFSYVQPSNMWISLQLDDI-----EIVRANNITPEYLFQNH	1299
Qy	1285	ITEDGRICTLAYRRW-VOSKAIPQLISL-LNNGRPLPPHSLTUTTRYDHDPE-QQIRQ	1344
Db	1299	LIDNYLCPALRRWGRQNNLLITEGVGLTLKNPMRQPPHLLTVVDVNYFSASEPQCHQ	1355
Qy	1342	QVFSDFGRLLOAAARHEAGMARQNEGDSLIN-----VO-HTENRWAVTGRTEYDNK	1399
Db	1359	TLAFSDGFGVLLSARVETGPSYFQPEPENGLLVDDKGNLVQLEVDQKQWVSGRTEYDNK	1411
Qy	1396	GOPIRTYQPYFLND--WRVYNSNDSARCEKAYADTHVYDPTGREIKVITAKGWFRRLTFT	1455
Db	1419	GLPFRYQPYFFDNWILYIANN--RTLKEAYADTHYDPLGREIKVITAKGLERTHYF	1470
Qy	1454	PWFTVNEDENDTAAEV 1469	
Db	1477	PWFVISEDNDTASEI 1492	
RESULT 12			
Qy	Q8D1P6	YERPE PRELIMINARY; PRT; 1516 AA.	
Id	Q8D1P6	YERPE PRELIMINARY; PRT; 1516 AA.	
AC	Q8D1P6;		
DT	01-MAR-2003 (TrEMBLrel. 23, Created)		
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)		
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)		
DE	Putative toxin subunit.		
GN	Name=tcaC1; OrderedLocusNames=y0185;		
OS	Yersinia peatis.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Yersinia.		
OX	NCBI_TaxID=632;		
RP	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=KIM5 / Biovar Mediaevalis;		
RX	MEDLINE=42137863; PubMed=12142430;		
RX	DOI=10.1128/JB.184.16.4601-4611.2002;		
RA	Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,		
RA	Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,		
RA	Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,		
RA	Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,		
RA	Perry R.D.;		
FT	"Genome sequence of Yersinia pestis KIM.;"		
RL	J. Bacteriol. 184:4601-4611(2002).		
RL	EMBL; AE013618; AAM83779.1; -; Genomic_DNA.		
DR	GO; GO:0005737; C:cyttoplasm; IEA.		
DR	GO; GO:0008305; C:integrin complex; IEA.		
DR	GO; GO:0007160; P:cell-matrix adhesion; IEA.		
DR	InterPro; IPR000413; Intergrin alpha.		
DR	InterPro; IPR004048; Reg_chr_condens.		
DR	InterPro; IPR003284; Sal_SpVb.		

DR	Pfam; PF01839; EG-GAP; 1.	Query Match	45.2%;	Score 3569.5;	DB 2;	Length 1516;
DR	Pfam; PF03534; SpvB; 1. <td>Best Local Similarity</td> <td>49.3%;</td> <td>Pred. No. 1.5e-218;</td> <td></td> <td></td>	Best Local Similarity	49.3%;	Pred. No. 1.5e-218;		
DR	PRINTS; PRO1341; SALSPVBPROT.	Matches 748;	Conservative 211;	Mismatches 486;	Indels	71; Gaps 29
DR	PROSITE; PS00626; R0C1.2; UNKNOWN 1.					
SQ	SEQUENCE 1516 AA; 170748 MW; 40DE58460A811CA7 CRC64;					
Qy		1	MONS-QDPSITELSLPKGGGAI	TGMEALTPTGPDMAALSILP	PISAGRGVAPFTLNY	59
Db		21	MENSKQQVAVAPLSLPKGGGAI	TGWDLSLGIPIGSGMATILP	PISAGRGVAPSLTILSY	80
Qy		60	NSGAGNSPFGLGWDCNVMT	IRRRTHFGVPHYDETDTFLG	PEGEVLVWADQPRD--ESTLQ	117
Db		81	SSGSGNGPFGLGWGLGTW	AIERRTNAQVPRYDEYDEFL	APNGEVMVVAADFGS	140
Qy		118	GINIGATFTVTGYSRRL	ESHRSRLYEWPKT-TGKT	DFWLIYSPDQGVHLLK	176
Db		141	SLN-GEQFSVTRYLPRI	EFGNPHRIEYWRPRTNS	QAPFWLHSSDQGHCLG	199
Qy		177	NPSQTTQTAOWLL	EASVSRGEQIYYOYRA	EDDTGCEADEI-----THL	231
Db		200	DPLHEHIAEWLL	BSVLSGEHIGYQQA	DEQIDEPSYKAEQNH	259
Qy		232	VYGNRTASETLPLGLD	SAPQADWLFYLVFDY	GERSNNLKTTPAFST	291
Db		260	VYGNRQAAVELYCLT	-QOPAPTSLWFLSLI	FDHGEYSNIAEQVPVI	318
Qy		292	RYEYGFETRRLCRQV	LMYHHLQALDSKIT	EHNGPTLVSRLLIN	351
Db		319	HFNYGFVETRLRCQV	LMYHNLGALGDB	PDQAQ-TLVSRLRLH	377
Qy		352	VGHEQDCNVVTLPLE	LAYQDFSRRHHA	WQPMVLANFNAIQR	411
Db		378	LAHEPDGTRKSLP	LEFDYQDFTRDAL	GWOPLTDWAEFN--YQ	435
Qy		412	QDKGAWYRSAQL	GEIG-SDAVTW	EKKQPLSVIPSL	470
Db		436	QDSGHWIYRPVR--	QPCADGII	FGAAQLPSLPAM	493
Qy		471	LRGVH-----SQ	RPDGSWTRFTPL	NALPVYTHPRAQL	524
Db		494	LAGYFSRDPDL	SRPDL	SWTQFIPLSLTP	553
Qy		525	YANTRDFAGK	QDVVQSGDITL	VPYGADPRKLVAF	584
Db		554	YTNLCDSFAA	ATQVAQDDITL	PLFGVHFTLNAF	613
Qy		585	GRGRFGOPITL	PGFSQPA	TENPAQVYLADL	644
Db		614	GHGRFGHPL	SLPGFNQV	PEQFNPLAIYLAD	733
Qy		645	PVTLRPFEG	LRFDHCTQL	QMAVQGLGVAS	704
Db		674	PLAIALPTG	IRFDNSCQL	SLADIQGLGVAS	733
Qy		705	NMGVHHTLYR	SSQFQWLDE	KAAALTTQGT	764
Db		734	NMGAESQLLY	RSVYQFV	MLDEKAQAKQGR	793
Qy		765	RYARGAWD	GRERERF	GGYVEQTS	824
Db		794	RYTFHGFYD	GVQREFS	GGRGVDTL	853
Qy		825	YWR-DDQAF	AGFS	PRFTTWQDNK--	881
Db		854	YWGDDQAY	SLPLRLTK	FINNTQ	913
Qy		882	NKHVPYV	TFPRSV	RRRLQHTDS--	939
Db		914	LATQPSV	NGSRVYQ	ROITQSSADG	973









SQ	SEQUENCE	1446 AA; 162070 MW; 8DBAAD65CDC8949 CRC64;
	Query Match	28.8%; Score 2274; DB 2; Length 1446;
	Best Local Similarity	35.7%; Pred. No. 8.7e-116;
	Matches 550; Conservative 220; Mismatches 582; Indels 188; Gaps 39;	
Qy	2	QNSQDFSTELSLPKGGAGTGMCEALTPTCPDGMALSLPLPSACRGVAPAPFTLVNS 61
Db	13	QSLQPAVATPL-LPKGGAGTQISKGWGSVGTSAASLEVALPISPGRGVAPALSLSYQS 71
Qy	62	GAGNSPGLGWDNCNMVTRRTHRGVPHYDFTDFTLGPGEGLVAVADQPRDE----- 113
Db	72	TSGNGVGLGNLNTSKVARASRGVPTVANDLLIFGPGDVCL---PERDSCALYSTQ 128
Qy	114	-STLQGINLGNFTVTGYRSRLESFHSFLEWQKTKTDFWLIYSPDGOVHLLGSPQ 172
Db	129	VSRYNHDLNATYQVRYFSRVEGAFLRIEHW-VDIADPGFWLIHGADSLNLYGRRTS 187
Qy	173	ARISNPSTTQTAQWLLLEASVSSRGEIYYQYRAEDDTGCCADEITHLQATAQRYLHV 232
Db	188	SRIADPADMKRVAEWLDESNNALGEHLIYKPEDHQGLAED---HPRNFRADQYLSRV 244
Qy	233	YYGNRTASETL-----PGLDGSAPSQADWLFYLDYGERSNNLKTPPAFTTGSWLCRQ 287
Db	245	RYGNAKAHYVLYLWQEDSLDGLL-----WHFDLIFDYDQRDRSEPPPEYDEQFTWVRS 299
Qy	288	DRFRSYGFBIRYRRCQVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESIASTLV 347
Db	300	DPHSSFAFGELGNLRCQVLMFHH---FPNELGE--APLLTRRLLEHLQTTLYGNLL 354
Qy	348	FVRVVGHEQ--DG---NVVTLPLELAYQDFSPRHHAHQPMQDVLNANAIORWQLVDL 401
Db	355	---SAASQAWGDGTDWRVDQPPQFYQYTDFS--LESGIYTPLEPMAGLNDQOYQLVDL 410
Qy	402	KGEGPLGLLYODKAWYRSARLGEIGSDAVTWKMQPLSVIPSLOS----NASLVDIN 457
Db	411	YGDGLPLGILYRDDKAWLYREPVRDTTGGADAVAYGACQPLPRIPTADSATPVQTLTDLT 470
Qy	458	GDGGLDWITGPGRLGVHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLVL 517
Db	471	GDGRLDWVVAQPGMAGFTTLNPDRSWSNYATFSAPFAFFHPQOGMADLVGDGLSLDALI 530
Qy	518	GPKSVRLYANTR-DGFAKGDVQSGDITLVPVPCADPKLVAFSDVLGSGOAHLVVEVSAT 576
Db	531	GPRSVRLYANRRADGFAAATDI PHDED-RLPLSDSSTELVAFSDLLGTGQOHLIRIHN 589
Qy	577	KVTCWPNLGRGRFQOP---ITLPGFSQPAFENPAQVYLADLDGSGPTDLIYVHTNRLDI 633
Db	590	EIRVWPNLGRGFGKQLFATLP-YTYEA--FDSSQVRLADLDGSGASDVLYLQADGFQV 646
Qy	634	FLNKSNGFAPBPVTLRPPEGLRPFHTCQLQWADVQGLGVASLIISVPHMSPHHW----RC 689
Db	647	FMNRGGGLAAPPDQRPWPEGVRYDFCQFSADVLLGLGFSLSVLTVPHMAPLHWSLYAA 706
Qy	690	DLTN--MKPMLLNEMNMVHHTLYRRSSQFWLDEKAAALTGTQTPVCVLPPIHTLW 747
Db	707	DRAGSVKPYLLKASDNNLGAAGEVSTRSSAQEWLDEKNELRAAGSVAVSELPPVHVWV 766
Qy	748	QTETEDBSGNKLVTLTRYARGAWDGREREPFRGFYVEQTDSDHQAQGNAPERTPPALTK 807
Db	767	ROTQMDKVTGNTLQLFRYRQGFYDPREREPRFGFGLLLQTDDETSSQ--DQEDFTAPVLNK 825
Qy	808	NWYATG-----LPVIDNALSTEYWDQDQAFAGFSRFTTWQ 843
Db	826	TWFHTGRYPARPCTDHRSDLLARLPGEHVLRLDAATQTEQ----- 867
Qy	844	DNKDVPILTPEDDNGRYFNFRALKGOLLKSELSEYGLDDSTNKHVPYVTVEFRSQVRLQHTD 903
Db	868	-----PITDADATLQEWARALSGSVLESEVFGLDASQRPVLYSTRSCRVLVRLQALS 922
Qy	904	SRYPLWSSV-----VESRNYHYE-RIASDPQCSQNTILSSDRFGOPLKQLSVQYPRRQOP 958
Db	923	AHRE--YASMLPLSLEVITYEABELELDPMCEHSLNLANWDRYGSTLHLSVSNYARRKKP 980

Qy	959	A-INLYPDTLDPKLLANSYDDQORQLRLTYQSSWHHLTNNVTVRVLGLPDSTRSDIFTYG 1017
Db	981	GDAPPPADPHQQQWWEASHDDAQOQFYLNEHBAIYLDSPQSWRLGLPYRTRGDAMLIP 1040
Qy	1018	AENVPAGLNLLELSDXNSLIADDKPREYLGOQKTAFT---DQONTTPTLOTPTTQALIAF 1074
Db	1041	ASALTPAQISVIEQPADPSGPPA-TLPRTLTSLSVQRYIGCGDGE-----ASFOALADA 1092
Qy	1075	TETTVFNQSTLSAPNGSIPSDKSLSTTLEOAGYQQTNYLFPRTGDKVVAHHGYDYDTA 1134
Db	1093	VETAELDHALSAYERVMDSVTLAEKLVEIGYQOMPSPFLPADSLN-LWSVKRGFATYAGQ 1151
Qy	1135	AQFWRPQKQSNLTQKITLIWDANYCVVQTRDAAGLTTSAKYDWRFLTPVQLTDINDN 1194
Db	1152	BHFPHTQFPRPTRSHGWSLVEYDAYHLFATRIITDPADCVTTAEYDVRVLQPKRIIDPNQN 1211
Qy	1195	QHLITLDALGRPITLRFWGTENGKMTGYSSPEKASFPSPSDVNAAIELKKPLPVAQCQVY 1254
Db	1212	QERADYDAFGRVWATSVFGTELGEAVGFPPLNRAGHYWASAGEVALQ----- 1258
Qy	1255	APESWMPVLQKTFNRLAEOQWQKLYNARIITEDGRICTLAYRRWVOSQKAIPLQILSLN 1314
Db	1259	-PE-----YALGRQ-----ASALYYDGN-----TVLGLVH 1282
Qy	1315	NGPRLPHSLTLTTDRYDHDPEQOIROOVVFSQDGFGRLLQAAARHEAGMARQNRNEDGSLI 1374
Db	1283	-----IPLATAVLVADRYPELDLDKQIRISMASIDGFGRTLQTRQKVEDGDAYSVDENWGLE 1338
Qy	1375	I-----NVQHTENRWAVTGRTEYDNKGQPIRTTQPYFLNDWRVYSNDSARQEKAYADTH 1429
Db	1339	LVDGKPKIVHASPRWRISERVEYNNKGLAVRVYRPFANSHLYVNDASIRSN--IVDKQ 1396
Qy	1430	VYDPIGREIKVITAKGWFRTLTPTPWETVNEDENDTAAEV 1469
Db	1397	FYDPLGRPTTITAKGWMRRQYRVVWYTISEDENDTAAEV 1436

Search completed: February 16, 2006, 21:42:55  
Job time : 205.739 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:46:54 ; Search time 9.48673 Seconds  
(without alignments)  
2208.600 Million cell updates/sec

Title: us-10-754-115-45

Perfect score: 7901

Sequence: 1 MNSQDPSITSLPKGGGA.....WFTVNEDENDTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107819 seqs, 14214640 residues

Total number of hits satisfying chosen parameters: 107819

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6031.5	76.3	1476	US-10-647-956A-4	Sequence 4, Appl
2	169.5	2.1	14130	US-11-175-689-9	Sequence 9, Appl
3	162.5	2.1	16990	US-11-175-689-7	Sequence 7, Appl
4	150	1.9	2228	US-11-124-367A-270	Sequence 270, Appl
5	150	1.9	3377	US-11-124-367A-271	Sequence 271, Appl
6	150	1.9	3841	US-11-124-367A-272	Sequence 272, Appl
7	139	1.8	2515	US-11-113-424-53	Sequence 53, Appl
8	136	1.7	2769	US-11-113-424-14	Sequence 14, Appl
9	131.5	1.7	2105	US-11-052-554A-173	Sequence 173, Appl
10	127.5	1.6	2204	US-11-052-554A-174	Sequence 174, Appl
11	127	1.6	2591	US-10-453-372-718	Sequence 718, Appl
12	127	1.6	2602	US-10-453-372-716	Sequence 716, Appl
13	127	1.6	2617	US-10-453-372-666	Sequence 666, Appl
14	127	1.6	2617	US-10-453-372-732	Sequence 732, Appl
15	127	1.6	2617	US-10-453-372-734	Sequence 734, Appl
16	127	1.6	2617	US-10-453-372-736	Sequence 736, Appl
17	127	1.6	2617	US-10-453-372-738	Sequence 738, Appl
18	127	1.6	2617	US-10-453-372-740	Sequence 740, Appl
19	127	1.6	2617	US-10-453-372-742	Sequence 742, Appl
20	127	1.6	2617	US-10-453-372-744	Sequence 744, Appl
21	127	1.6	2617	US-10-453-372-746	Sequence 746, Appl
22	127	1.6	2617	US-10-453-372-748	Sequence 748, Appl
23	127	1.6	2617	US-10-453-372-750	Sequence 750, Appl
24	125.5	1.6	1579	US-11-052-554A-9	Sequence 9, Appl
25	125	1.6	1323	US-10-517-939-312	Sequence 312, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-647-956A-4  
; Sequence 4, Application US/10647956A  
; Publication No. US20050251878A1

GENERAL INFORMATION:

; APPLICANT: french-Constant, Richard

; APPLICANT: Bowen, David

; APPLICANT: Rocheleau, Thomas

; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: 61645

; CURRENT APPLICATION NUMBER: US/10/647,956A

; CURRENT FILING DATE: 2003-08-26

; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514

; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26

; PRIOR APPLICATION NUMBER: US 60/191806

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 8

; SOFTWARE: Patent in version 3.0

; SEQ ID NO 4

; LENGTH: 1476

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-10-647-956A-4

Query Match 76.3%; Score 6031.5; DB 6; Length 1476;

Best Local Similarity 75.9%; Pred. No. 0;

Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

QY 1 MNSQDPSITSLPKGGGATGCGEALTPTGPDGMAALSPLPISAGRGVAPATLNN 60

DB 1 MNSQTSVTSLSLPKGGGATGCGEALTPTGPDGMAALSPLPISAGRGVAPATLNN 60

QY 61 SGAGNSPPLGMDNCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

DB 61 SGTGNPPPLGMDNCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVAALNAGQADIRSES 120

QY 115 TLOGINLGATFTVGYRSRLESFSLRYWQPKTKGDTFWLIYSPDQGVHLLGKSPQAR 174

DB 115 TLOGINLGATFTVGYRSRLESFSLRYWQPKTKGDTFWLIYSPDQGVHLLGKSPQAR 174

QY 175 ISNPSQTTQAWLLEASVSRGQIYYQYRAEDDTGCEADEITHLQATQRYLHIVY 234

DB 175 ISNPSQTTQAWLLEASVSRGQIYYQYRAEDDTGCEADEITHLQATQRYLHIVY 234

QY 181 ISNPLNVTQAWLLEASISSEBQIYYQYRAEDGAGCETDELAHPATVORYLOTVHY 240

DB 181 ISNPLNVTQAWLLEASISSEBQIYYQYRAEDGAGCETDELAHPATVORYLOTVHY 240

QY 235 GNRITASETLPLGDSAPSOADWLFYLVFDYGRSNLKTTPAFSTTGSWLCRODRFSRYE 294

DB 235 GNRITASETLPLGDSAPSOADWLFYLVFDYGRSNLKTTPAFSTTGSWLCRODRFSRYE 294

QY 241 GNLTSADVFTPLNGDDPLKSGWMFCLVDYGERKNLSLSEMPFLKATGNWLCRKDRFSRYE 300

DB 241 GNLTSADVFTPLNGDDPLKSGWMFCLVDYGERKNLSLSEMPFLKATGNWLCRKDRFSRYE 300

QY 372, Appl

Sequence 8, Appl

Sequence 96, Appl

Sequence 624, Appl

Sequence 51, Appl

Sequence 72, Appl

Sequence 145, Appl

Sequence 115, Appl

Sequence 198, Appl

Sequence 314, Appl

Sequence 113, Appl

Sequence 73, Appl

Sequence 77, Appl

Sequence 250, Appl

Sequence 6, Appl

Sequence 218, Appl

Sequence 15, Appl

Sequence 15, Appl

Sequence 15, Appl

Sequence 15, Appl



Db 5510 L-----FSVMVSLAALWRW--AGITPDAVIG-----HSQ 5536  
Qy 426 GEIGSDAVTEKMOB-LSVIPSLQSNASLVINGDQOL-----DWIT 467  
Db 5537 CEIAAAHVAGALTPEAAAVVALRSV-LTDLAGAGAWASVLSPEEPTQLLARDWGKIT 5595  
Qy 468 GPLGRYHSQRDPDSWTRFTPL-----NALPVEY-----THPRAQLAD----- 505  
Db 5596 VAAVNGPASAVVSGDTTITELLITCEHENIDARAIPVDYPSHSPYMEHIRHQFLDELPE 5655  
Qy 506 -----LMAGLSDLAVL 516  
Db 5656 LTPRPSTIAMYSTVDGEPHDTAYDTTMTADYWRNIRNTRFHTDVAALIGAGEQVLE 5715  
Qy 517 IGPKSVRLYANTRDGFAGKDVQVQ-----SGDITLPVPGADPRKLVAFSVLSGQAHLEVEV 573  
Db 5716 LSPHPVLTOALT-----DTVEQAGGGAAPALRKDRDPAFAAALQQLHCHGISP 5767  
Qy 574 SATKVTCPNIGRGRFGQPIITLPGS-----OPATFNPAQVYLAD-----LOGSGPTD 622  
Db 5768 SNNVLYCQ-----ARPLTPTVAFQHQRYWLLPTAGDFSGANTHAMHPLDLTA--TE 5817  
Qy 623 LI-----YHTNRL-----DIPLNK-----SGNGFAEPTVILRPPGELRFD-----H 658  
Db 5818 LAENRGWFTGRIISPTQFWLNEHAVESAVLPFGTGYE-LALHVADRAGYSVNNELIVH 5876  
Qy 659 T-----COLOMADVOGLGVASLILSVPMSPH--HWRCDLTNMKP-WLLNEMNN 704  
Db 5877 TPLLAGHDTADLQITVTDMDGROSINI--HSRPHIGHDNTTGTDEQPEWVLH----- 5929  
Qy 705 NMGVHTLRLYRSSQFWLDEKAAALTTCQT-----PVCYLPFPPIHTLWOTETED----- 753  
Db 5930 -----ASAVLTAQTDDHNLPLTPVWPPPPGTAIEVDYDD 5968  
Qy 754 EISGNKLVTLRYARGAWG-----RERFRFGY----- 783  
Db 5969 AAQGYNTGPTFGQVQRIWRDHATPDVIVAEVLPEDTIDGIGHPALFDAAHLPLAL 6028  
Qy 784 ---VEQTDHQAQGNAPERTPPALT-----KNWYATGL 814  
Db 6029 QPPTNDTDDTWTADTGDQVRLPYAFTGISLHATHATRLVRLTRTGADAITVHTSDTGA 6088  
Qy 815 PV--IDNALSTERYWRDQAFAGSPRFT-----TWQDNKDVPILTPEDDSRYWFNRL 865  
Db 6089 FVAIIDSLLIT--RPLTTATGSAPTAAGLLHLSWPPHPTDITDITDITD-----AL 6138  
Qy 866 KGOLLRSE-----LYGLDSTNKHVPYTFERSQVRLQHTDSRYPLWSSVWESRN 918  
Db 6139 RYQVIAEPTQPLRYLHDLHTSTDLHTTTEADV-----VWMPVPVPSNE 6183  
Qy 919 YHYERIASDPOCS-----QNITLSSD-----RFG-----QPLKQL 948  
Db 6184 ELQAHQASDTAVSSRIHTLRTQLTLVQDNLHTPTTGTTRLVIVTRHGVSATSDPVDL 6243  
Qy 949 S-----VOYPRRQQA-----INLYPDTLPDKLLANSYDDQQRQLRLTYQOQSWH 993  
Db 6244 AHAAMVGLIRSAQNEHPCREFTLLDDTNDNSDTLTALTLPRENQ-----LAIRRDIT 6298  
Qy 994 -----HUTNNVRLVGLPDSRSDIFTYG-----AEN-VPAGGLNLELLSDKNSLI 1038  
Db 6299 IPRLTRHSDGALTAPVVVDPEGTGLITGTTGALFAELVSAHGVRLHLLTSR----- 6355  
Qy 1039 ADDPREVLGQKATYD-GQNTTLPQTPTRQALIAFTETTVENQSTLSAFNGSIPSDKL 1097  
Db 6356 -----GPOHAGATLQORLTDL-----GAHTITACDISDEALALVNSVPTQHR 6401  
Qy 1098 STTL-----BQAGYQQTNYLFPRTGEDKVVVAHHGVHTDYGTAA--QFWRPQK 1142  
Db 6402 LTAVVHTAAVLADTPVTETLGTQDQVLAPKI--DAAMQLQLTYEHLNSAFIMF----- 6454  
Qy 1143 QNTQLTKITLIWANDYCVVQVTRDAAGLTTSYAKYDWRFTFPVQLTDINDNQHLITLDA 1202  
Db 6455 ---SSMAGMIGSPGQGNYA-----AANTALDALADYR-----HR 6485

Qy 1203 LGREITLRFWG-----TENGMKTYSSPEKASFPSPSDVNAALIEL 1242  
Db 6486 LGLPATSLANGYQWTHGLTAHLTDVLDLARMTLGLMPIATSHGLALFDAALATGQPVSI 6545  
Qy 1243 KKPLPVAQCQVYAPESWM-PVLS-----OKTFNRLAEQDWOKLYNARI 1284  
Db 6546 PAPINTHTLARHARDNTLAPILSALITTPRRRAASAAATDLAARLNGLSPOOQOQTLATLV 6605  
Qy 1285 ITEDGR1-----CTLAYRRWQSKAIPQLISLNN-GPRLPPHSLTLTTDRYD 1332  
Db 6606 AAATATVLGHTTPESIPATAPKDLGIDSLTALERNTLTHNTGLDLEP-----TLIFD 6659  
Qy 1333 HDPEQIIRQQVVS-DGFRLLQAAARHEAGMARQNEDEGSLII-----NVQHTENR 1383  
Db 6660 HPTPFAVAEHLLEQIPGIGALVPAPVPIAAG-----RTEEPVAVVMACRFPFGVASADQL 6715  
Qy 1384 W--AVTGRTEYDN-----KGQPI-----RTYQPY--FLND 1409  
Db 6716 WDLVIAGRDVVGNFPADRGMDVEGLFDPDPAVGKTYTRYGAFLLDD 6761

## RESULT 3

US-11-175-689-7

; Sequence 7, Application US/11175689

; Publication No. US20060024806A1

; GENERAL INFORMATION:

; APPLICANT: STINEAR, TIMOTHY P.

; APPLICANT: COLE, STEWART T.

; APPLICANT: LEADLAY, PETER F.

; APPLICANT: SMALL, PAMELA L.C.

; APPLICANT: JOHNSON, PAUL D.R.

; APPLICANT: JENKIN, GRANT A.

; APPLICANT: DAVIES, JOHN K.

; APPLICANT: HAYDOCK, STEPHEN F.

; TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING

; FILE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES

; FILE REFERENCE: 03495.0329-01

; CURRENT APPLICATION NUMBER: US/11/175,689

; PRIOR FILING DATE: 2005-07-07

; PRIOR FILING DATE: 2004-11-15

; PRIOR APPLICATION NUMBER: 10/987,592

; PRIOR FILING DATE: 2003-11-14

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: Patent In Ver. 3.3

; SEQ ID NO 7

; LENGTH: 16990

; TYPE: PRT

; ORGANISM: Mycobacterium ulcerans

US-11-175-689-7

Query Match 2.1%; Score 162.5; DB 7; Length 16990;

Best Local Similarity 18.5%; Pred. No. 0.034;

Matches 342; Conservative 186; Mismatches 631; Indels 685; Gaps 86;

Qy 4 SQDPSITSLPKKGGGAIKMGEGALTPTG-----PDGMAA-----LSLPLPISAGRYA-- 52

Db 11465 NQDQASGLTAPHGPSQQRVINQALANAGLTHDQDVAEANGTGTGLGDPLEASALHATY 11524

Qy 53 -----PAPTLNYSAGNSPFGLGWDCNVMTIRRTTHFGVP---HYDETDFLG-PE 100

Db 11525 GHHTTPOPLWGLSIKSNIGHTQAAAGAAAGVVMQIATHTATLTPATLHVDQSPHIDWSS 11584

Qy 101 GEVLVV-----ADOPRDESLTGILNGLATFTVTGYSRLESFSLRYWPKTKTKT 152

Db 11585 GTVRELLTEPIQWPNTHPRTAAV-----SSFGISGNTNAHLIQ-----QPTPTDT 11630

Qy 153 DFWLIYSPD---GQVHLGLKSPQARISNPSTOTTAQWLLA-----SVSSRGEQIYYQYR 205

Db 11631 Q-----TPNTTGTGSDPAVGSDPAVG-----LVWPLSARSPAGLSAQAARLYOHL 11676

Qy 206 AEDDTGCEADEITHHQAQRYLH-----IVVY--GNRTASETLPLGLDGSAPSQADW 256

Db 11677 AHPD---LDPIDVAHSLATRRSHHPRATITTSIEHSHNNHDTTDAALHALANNGTHP 11734  
 Qy 257 LFYLVDFYDGRSNNLKTTP-----AFSTTGWLCRQDRSRVEY---GFEIRTRLCR 306  
 Db 11735 LL-----SRGLLTPQSGKTVFVPPGQS-----QYPMGAD-----LYR 11769  
 Qy 307 QVLMY-HHLQALDSKITEHNGPTLVSLRLILNYDESATLSTLVFVRVGHGBOGNNVTLP 365  
 Db 11770 QFPVFAHALDACDAALQFTGWSVLAVL---HDEPEAPS-----LERV-----VQPV 11815  
 Qy 366 LELAYQDFSPRHHAHQPMVDLANFNATQRQLVDLKGEGPLGLLYQDKGAWYRSQRL 425  
 Db 11816 L-----FSWVLSLAALRW--AGITPDAVIG-----HSQ 11842  
 Qy 426 GEIGSDAVTWKMP--LSVIPSLQSNASLVINDGGQL-----DWTIT 467  
 Db 11843 GEIAAAHVAGALTLPAAAAVVALRSV--LITLAGAGAMASVLSPEEPLTQLLARWDGKIT 11901  
 Qy 468 GPLRGVHSQRPDGSWTFTEL-----NALPVEY-----THPRAQLAD--- 505  
 Db 11902 VAAVNGPASVVSGBDTTAITELLITCBHENIDARAIPVDVPSHSPYMEHIRHQFLDELPE 11961  
 Qy 506 -----LMGAGLSDLVL 516  
 Db 11962 LTPRPSTIAMYSTVDGEPHDTAYDTTWTADYWNINRTVRPHDTVAALLGAGEQVPLE 12021  
 Qy 517 IGPKSVRLYANTRDGFAGKDVQVQ---SGDITLVPVGDPRKIVAFSDVLGSGQAHILVEV 573  
 Db 12022 LSPHPVLTOAIT-----DTVEQAGGGGAAPALRKRDPDAVAFAAALQLHCHGISP 12073  
 Qy 574 SATKVTWPNLGRGFGPITLPFGS-----QPATENPAQVILAD---LDGSGPTD 622  
 Db 12074 SWNVLYCO-----ARPLFTYAFQHORYWLLPTAGDFSGANTHAMHPLDITA--TE 12123  
 Qy 623 LI-----VYHTNRL-----DIFLNK-----SGNGFAEPVTLRFPFGLRFDHTCOLQMA 665  
 Db 12124 LAENRGWFTGRIISPRQPLWNEHAVESAVLFPGTGFVE-----LALHVA 12168  
 Qy 666 DVQGL-GVASILSVPHMSPHHWRCDL-----TNMKPWLNNEMNNMNGVHHTLRYRSS 717  
 Db 12169 DRAGYSSVNELIVHTPLLLAGHDADLIQITVTDGDMGROSLNTHSPHIGHDNTTTGDE 12228  
 Qy 718 SQFWLDEKAAALTGTQPCVCLPFPPIHLMQTEDEISGNKLVITLRYARGAWDGRERE 777  
 Db 12229 QPEWVLHASAVLTATQTDHNLPL----- 12252  
 Qy 778 FRGFGYVEQTDHQLAQNAPERTPPALTKNMYATGLFVIDNALSTBYWRDQAFAGFS- 836  
 Db 12253 -----TPVPWPPCTA-----ALEVDDFYDDLAAQGYNY 12281  
 Qy 837 -PRF-----TWQDN-----KDVPLTPEDDNSRYWPNRALKGOLLSELYGLDDSTNK- 883  
 Db 12282 GPTFGQVQIRWRDHATPDVIAEVELPDTDIDGYIHPALFDAALHPALLALTQPTNDT 12341  
 Qy 884 -----HVPYTVTEP-----RSQVERLO-----HTD--SRYPVLNWSV 913  
 Db 12342 DDTWTADGQVRLPYFATGILHATHATRLURVLRTRTGADAITVHTSDITGAPV---AI 12398  
 Qy 914 VESRNYHYERIASDPQS-----QNITLSSDRFGOPLKQLSVQY 952  
 Db 12399 IDS-----LITPLTITATGASAPATTAAGLLHLSWPPHPTTDTTDDTALRYQVIAE 12451  
 Qy 953 PRROOPA-----INLYPDTLP--DKLLANSYDDQQRQLRLTYQQSSWHHLTN 997  
 Db 12452 PTQQLPYRLDLHTSTTEADVVMVFPVPSNEELQAQASD-----TAVSSRIHTLTR 12504  
 Qy 998 NTVEV-----LGLPDS-----TRSDIFTYGAEVNP-----AGGNLELSDKNSLIA 1039  
 Db 12505 QTLTVQDQWLTHPTDTGTGLVIVTRHGVSSTAHPDPVDPDLAAHAWGL---IRSAQN--- 12557  
 Qy 1040 DDKPREVLGQOKTAYTDGQN-----TTPLOTPTFRQALIAFTETTVF-----NOSTLSAFN 1089

Db 12558 -----EHPGRFTLLDDTNTNSDTLTATTALTPRTENQAIARRDTHIPLRTHSSDGALT 12612  
 Qy 1090 GSIPSDKSLTLEAGYQQOYNILFPRTGEDKVVAAH-----HGYTDYG----- 1132  
 Db 12613 APVVVDEPGVTLTGTGTGTLGALFAEHLVSAHGVRLHLLTSRRGPOAHGATDLQOURLTDL 12672  
 Qy 1133 -----TAAQFWRPQKOSN-----TQUTG-----K 1151  
 Db 12673 GAHVTITACDISPEALAAALVNSVPTQHRILTAVVHTAAVLADTPVTLTGQDLDOVLAPK 12732  
 Qy 1152 ITLIW-----DANYCVVVQTRDAAGLTTSYAKYDNRFLFPVQLTIDNNOHLITLDALG 1204  
 Db 12733 IDAAWQLHQLTYEHLNSAFIMFSGMAGMIGSPGCGNYAAANTALDALADYRH-----RLG 12787  
 Qy 1205 RPTLRPWG-----TENGMTCYSSPEKASFSPSPDVNAAIELKK 1244  
 Db 12788 LPATSLAWGYWQHTGTHTAHLTDVLDLARMTRLGLMPTATSHGLALFDAALATGQPVSIIPA 12847  
 Qy 1245 PLPVAQCQVYAPESWM--PVLS-----OQTFNRLAEQDQWKLYNARIIT 1286  
 Db 12848 PINHTILARHARDNTLAPILSALITTPRRRAASAATDLAARLNGLSPOQOQOTLATLVAA 12907  
 Qy 1287 EDGRI-----CTLAYRWVQSKAIPOLISLNN--GPRLPPhSLTLTTTDRYDHD 1334  
 Db 12908 ATATVLGHHTPESISPATAFKDLGIDSLTALRLNTLTHNTGLDLP-----TLIFDHP 12961  
 Qy 1335 PEQIIRQOVVFS--DGFRLILQAAARHEAGMARQNEGSLII-----NVQHTENRW- 1384  
 Db 12962 TPHAVABHLLLEQIPGIGALVPAPVIAAG---RTEEPVAVVGMACRFPGGVASADQLWD 13017  
 Qy 1385 -AVTGRTEYDN---KGQPI-----RTYQPY--FLND 1409  
 Db 13018 LVIAGRDVVGNFADRGWDVEGLFDDPDPDVGKTYTYGAFLLD 13061

RESULT 4  
 US-11-124-367A-270  
 ; Sequence 270, Application US/11124367A  
 ; Publication No. US20060024700A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Michele Cargill  
 ; APPLICANT: Hongjin Huang  
 ; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
 ; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof  
 ; FILE REFERENCE: CL001519. ORD  
 ; CURRENT APPLICATION NUMBER: US/11/124,367A  
 ; CURRENT FILING DATE: 2005-05-09  
 ; PRIOR APPLICATION NUMBER: US 60/568,846  
 ; PRIOR FILING DATE: 2004-05-07  
 ; PRIOR APPLICATION NUMBER: US 60/582,609  
 ; PRIOR FILING DATE: 2004-06-25  
 ; PRIOR APPLICATION NUMBER: US 60/599,554  
 ; PRIOR FILING DATE: 2004-08-09  
 ; NUMBER OF SEQ ID NOS: 34460  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 270  
 ; LENGTH: 2228  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-11-124-367A-270

Query Match 1.9%; Score 150; DB 7; Length 2228;  
 Best Local Similarity 17.5%; Pred. No. 0.013;  
 Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;  
 Qy 220 HLOATAQRYLHVYVYGNRTASSETPLGDLGDSAPQADWLFYLVDFYDGRSNNLKTTPAFST 279  
 Db 366 HLKFTQDVYSAV--KENSTEATLAVITACNPINEPLFYHILNPDRRFKISRTSGVLST 424  
 Qy 280 TGSWLCHQDRFSYEVGFETRLCRQVLMYHHLQALD--SKITEHNGPTLVSRLLNY 337  
 Db 425 TGTPDFDEQO-----EAFDVVVEVTEEHKPSAVAHVV--- 456

Qy	338	DESIASTLVPVRRRGHEQDGVN---VTLPPLLEYAQD-----	376
Dd	457	-----VKIVIVEDQNDAFVNFLPYAVVVKVDTEVGHVIRYVTVADRDGRNGE	505
Qy	373	---PSPRHHHQWDMVLANFNATORMQLVDLKCEGLPGLLYQDKGAWYRSQAURLGEIG	429
Dd	506	VHYLYKGGHHEHFQ-IGPLGEISLKKQFELDTLNKEYLVTVVAKOG-	550
Qy	430	SDAVTWEKMOPLSVI-----PSLQSNASILDINGDCQLDWMVITGPCLR	472
Dd	551	NPAFSAEVPIVTVMNKAMPVFEPFYSAETAESIQVHSPPVVHYQAN-----SPEGLK	603
Qy	473	GYSHQRPDGSWTRFT-----PLNAL-PVEY-THPPRAQLA-----DLMGA-	509
Dd	604	VFSITDDGPFQSQTINFNTGVINVIAPLDPEAHPAYKLSTRATDSLTGAHAEFVFDIIV	663
Qy	510	-----GLSLDLVLIGPKSVRLYANTRD-----	530
Dd	664	DDINDNPVFAQQSYAVTLSEASVIGTSVQVRATDSDBENRGISYQMFGNHKSJDHP	723
Qy	531	-----GFAGKGDVVQSGDITLTPVPGADPRKLVAFSDVLGSCQAHLEVSATKVTCW	581
Dd	724	HVDSSTGLISLRITLDYEQRQHTIFRVADVGMPTLSSDV-----IVTVDTDLNDN	776
Qy	582	PNLGRGFPQGITLPGPSQPATE---FNPAQVYLAD-----LDG-----	617
Dd	777	PPL----FEQOIYEARISEAPHGHFVTCVKAVDADSSDIKLOYLSILSGNDHKHFVIDS	832
Qy	618	-SGPTDLIYVHTNFLDIPLFKSNGNGPAEPVTLRFPEGLRPHDTCOLQWADVQGLVASLI	676
Dd	833	ATGIITLSNRHRHALKPFS-----LNLVSVDGV-FRSSTQVHVTVTGG-----N	876
Qy	677	LSPVPHMSPHWRCDLTNNKP--WLLNEWMNNMG-----VHHTLYRSSSQFWLDEKA	726
Dd	877	LHSPAFLONEYVELAEANPLHTLMVEVKITDGSGLYGHVTHYI VDFPAKORYFINERG	936
Qy	727	AALUTT-----GQFVCYCLPFPFIHLTMQTEDEISGNKLVT 763	
Dd	937	QIFTLEKLDRETPAEKVVISVRLMAKDAGGKVAFC-----TVNVILTDDNDNAPQFRAT	989
Qy	764	LRY-----ARG-----AMDGREBERFGYGYEQTDSSHOLAQNAPERTPALTKN	808
Dd	990	-KYEVNIGSSAAKGTSVVKVLASDADBGSNADITYAIEADSES VKENLEINKLSGVITTK	1048
Qy	809	WYATGLP-----VIDNA-----LSTEYMRDDQAFAGFSRFTTWQDNKDVP 850	
Dd	1049	ESLIGLENEFFTFVRAVDNGSPSKSVLVVYVKILPEMQLPKFSEFFYFTFVEDVPI	1108
Qy	851	TPEDDSRYWFNRALKQGLLRSELGYLDDSTNKHVPYTVTFERSQV---RRLQHTDSRYP	907
Dd	1109	GTEIDLIR----AEHSGTGLVSLVKGNTPESNRDES FVIDRQSGRLKLEKSLDHETT KW-	1163
Qy	908	VLSSVVESRNYHYERTASDPQCSQNTLSSDRGQPLKQLSVQYPRQOQAINLYP---	964
Dd	1164	-----YQFSILARCTQDDHEWVASVD-----VSIQVXANDNGSPVFESSPYEA	1206
Qy	965	---DTLPDKLLANSYDQORQLRTYQQSSWHLLTNNTVRLGLPDSTRSDIFYGAENV	1021
Dd	1207	FIVENLPG-----GSRVQIRASDADSG-----TNGQV-MYSLDQSQSVIESFAIM	1254
Qy	1022	PAGGLANL---ELLSDKN-----SLIADKPREYLGQOKTAYTDGQNTTLPOTPTRQALIAF	1074
Dd	1255	ETGWITTLTELKDHEKRNYQIKVVASDH-GEKIQLSSTAIVDTVTVDVNDSPPR----	1307
Qy	1075	TETTVFNQSTLSAPNGSIPSDK-----LSTT---LEQAGYQOOTHYLFPRTGEDKWV	1123
Dd	1308	-----FTAIEIKGTGVSEDDPOGGVIALISTTTDADSEENRQVTFYI---TGGDPL--	1354
Qy	1124	AHHGYTDYTAAQFWR-----POKOSNTQLAGKITLIWDANYCVVQTRDAAGLTTS	1175
Dd	1355	---GQFAVETIQNEWKYVVKPLDREKRDNVLLT--IT-----ATD 1390	
Qy	1176	AKYDWRELTPVQLTIDINDNOHL-----ITLDALGRPITLFWGTE-----NGRWMT	1221

```

Db      1391 GTFSSKAIVEVKYLDANDNSPVCEKTYLSDTIPEDVLPGLKINQISATDADIRSNAEIT- 1449
Qy      1222 Y----SSPEKASPSPPSDVNAALTEKKLPVPAQCQVYAPESWMPVLSQKTFNRLAEQDQW 1277
Db      1450 YTLGSGAEKFLNPDTG-----ELKTSFPL-----DREQ 1480
Qy      1278 KLYNARIITED--GRICTLAYRRWGSQKAIPOLISLNL-NGPRL--PPHSL----- 1324
Db      1481 AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPEFSADPVAVTFENTEP 1530
Qy      1325 -TLTTRYDHPDQRIQRQVWFS-----DG-----FGRLLQNA-----A 1357
Db      1531 GTLLTRVQATDAGLNKILYSLIDSADQGSINELSGIIQLEKPLDRELOQAVYTLSLK 1590
Qy      1358 RHEAGMARORNEGSLIINVOHTENRWAIVTGRTEYDNKQPIRTYQPY 1405
Db      1591 AVDQGLPRRLTAGTVIVSVLDI-----NDNPPVFEYREY 1625

RESULT 5
US-11-124-367A-271
; Sequence 271, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 271
; LENGTH: 3377
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-271

Query Match      1.9%; Score 150; DB 7; Length 3377;
Best Local Similarity 17.5%; Pred. No. 0.025;
Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;

Qy      220 HLQATQRVLHIVYGNRTASETLPLGDSAPSQADMLFVLFVFDYGERSNNLKTPPAFST 279
Db      1976 HLKFTQDVTSYVV-KENSTEAETLAVITAGNPINEPLFYHINPDRRFKISRTSGVLST 2034
Qy      280 TGSWLRCQDRFSRYEGYGFETRRLRCQVLMYHHQLALD--SKITEHNGPTLVSRILNY 337
Db      2035 TGFPEFREQQ-----EAFDVVVEVTEEHKPSAVAHVV--- 2066
Qy      338 DESAIASTLVFVRVGHGQGNV---VTLPPELAYQD----- 372
Db      2067 -----VKVIVEDQNDNAPFVNLPFYAVVVKVDTEVGHVIRYTVTAVRDSDGRNGE 2115
Qy      373 ---FSPRHHAHQPMVDLANFNALQRWLVDLKGELPGLLVQDKGAWVRSQRGLGETG 429
Db      2116 VHYLYLKEHHEHFQ-IGFLGEISLKKQFELDTLNKEYLVTVVAKDG-----G 2160
Qy      430 SDAVTWEKKOPLSVI-----PSLQSNASLVNDINGDGLDWITGFLGR 472
Db      2161 NPAFSAEIVPTVMNKAMPVEKPFYSIAEISIQVHSPVVHVQAN-----SPEGLK 2213
Qy      473 GYHSQRPDGSWTRFT-----PLNAL-PVEY-THPRAQLA-----DLMGA----- 509
Db      2214 VFSYITDGPSPSOFTTNEFTGVINVTAPLDFFAHPAYKLSIRATDSLTGHAHEVFVDIIV 2273

```

Qy 510 -----GLSDVLIGPKSVRLYANTRD----- 530  
Db 2274 DDINDNPPVFAQQSYAVTLSEASVIGTSVQVVRATSDSPNRGSIQYMGFNHSHKSHDHF 2333  
Qy 531 -----GPAKGQVQSGDITLPVPGADPRKLVAFSDVLSGQAHLEVSATKVTWCW 581  
Db 2334 HVDSSSTGLISLRLTLDYEQSRQHTIFVRAVDGGMPTLSSDV-----IVTVDTDLNDN 2386  
Qy 582 PNLGRGRFGQPIITLPGFSQAPATE---FNPAQVYLAD-----LDG----- 617  
Db 2387 PPL-----FEQOIYEARISEHAPGHGFTVCYKAYDADSSDIDKLOYSTLSGNDHKGHFVIDS 2442  
Qy 618 -SGPTDLIYVHTNRDLDFLNKSGNGFAPVTLRPEGLRFDHTCQLOMADVQGLGVASLI 676  
Db 2443 ATGIITLSNLRHAKFPYS-----LNLVSVDGV-FRASSQVHVTVIGG-----N 2486  
Qy 677 LSVPHMSPHWRCDLTNNKP---WLLNEMNNMG-----VHHTLRYRSSSQFWLDEKA 726  
Db 2487 LHSPAPLQNEVEVELAENAPLHVLMEVKTTDGDGSIYGHVYTHVNDFAKDRFYINERG 2546  
Qy 727 AALT---GQTPCYLPPPIHTLMTQTEDEISGNKLVTT 763  
Db 2547 Q1FTLEKLDRETAEKVISVRLMAKAGKVAFC-----TVNVILTDDNDNAPQPRAT 2599  
Qy 764 LRY-----ARG-----AWDGRERFRGFGVQSDSHQLAQGNAPERTPPALTKN 808  
Db 2600 -KYEVNIGSSAAKTSVVKVLASDADSGNADITYAIEADSESVKENLEINKULSGVITTK 2658  
Qy 809 WYATGLP-----VIDNA-----LSTEYWRDDQAFAGFSFRFTTWQDNKDVP 850  
Db 2659 ESLLIGLENEPFTFFVRAVDNGSPSKESVVLVYVVKILPEMQLPKFSPPFTFTVSEDVPI 2718  
Qy 851 TPEDNDRYFNWRAKQGLRLSELYGLDDSTNKHKVPVTVTEFRSQV---RRLQHTDRYP 907  
Db 2719 GTEIDLIR---AEHSGTVLSLVKGNTPESNRDESFVIDRQSGRLKLEKSLDHETTKW- 2773  
Qy 908 VLWSSVSRNHYERTASDPQCSQNTLTASDRFGQPLKQLSVQYPRRQQAINALXP--- 964  
Db 2774 -----YQFSILARCTQDDHEWVASVD-----VSIQVANDNSVPFESSPYEA 2816  
Qy 965 ---DTPDKLIANSYDQQRRLTYQOSSMHHLTNNTVRLVGLPSTDRSDIFTYGAENV 1021  
Db 2817 FIVENLFG-----GSRVLIQIRASDADSG---TNGQV-MYSLDQSQSVIEBSFALNM 2864  
Qy 1022 PAGGLNL---ELLSKN-----SLIADDKPRYLQOQKTYTDGQNTTPTQTRQALIAF 1074  
Db 2865 ETGWITTLKELDHKEDKRDYQIKVVASDH-GEKIQLSSTAIYDVTVTVTVNDSPPR----- 2917  
Qy 1075 TETTVFNQSTLSAFNGSIPSDK-----LSTT---LEQAGYQOQTYNLPRTGEDKVVW 1123  
Db 2918 -----FTABIYKGTVSEDDPGGVAILSTTDADSEENRQVTFI---TGQDPL-- 2964  
Qy 1124 AHGYTDYGTAAQFWR-----POKOSNTQLTKITLIWDANYCVVQVOTRDAAGLTTS 1175  
Db 2965 ---GQPAVEITONEKVVYKPLDREKDNVLLT---IT-----ATD 3000  
Qy 1176 AKYDWRFLTPVLQDITDINDNHL-----ITLDALGRPITLRFWQTE-----NGKWTG 1221  
Db 3001 GTFSSKAIVEVKVLADNDSNPCEKLTYSDTIPEDVLPGLKIMQISATDADIRSNABIT- 3059  
Qy 1222 Y-----SSPEKASFPSPDNNAAIELKPLPVAQCVQVAPESMWPVLQKTFNRLAEQDWQ 1277  
Db 3060 YTLGSGAEKPKLNPDTG-----ELKTSTPL-----DREBQ 3090  
Qy 1278 KLYNARIITED--GRICTLAYRRVQSQKATPOLISLIN-NGPRL---PPHSL----- 1324  
Db 3091 AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPFSADPAIVTFENTEP 3140  
Qy 1325 -TLFTDRYHDPEQOIRQOVVFS-----DG-----FGRLLQAA-----A 1357  
Db 3141 GTLLTRVQATDADAGLNKILYSLDADGQFSINELSGIITOLEKPLDLRELQAVYVTLSLK 3200

Qy 1358 RHEAGMARQORNBGSLIINVOHTENRWAVTGRTEYDNKGOPIRTYQPY 1405  
Db 3201 AVQQLPRRLTATGTIVSVLDI-----NDNPPVEYREY 3235  
RESULT 6  
US-11-124-367A-272  
; Sequence 272, Application US/11124367A  
; Publication No. US20060024700A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: Hongjin Huang  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof  
; FILE REFERENCE: CL001519.ORD  
; CURRENT APPLICATION NUMBER: US/11/124,367A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,846  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/582,609  
; PRIOR FILING DATE: 2004-06-25  
; PRIOR APPLICATION NUMBER: US 60/599,554  
; PRIOR FILING DATE: 2004-08-09  
; NUMBER OF SEQ ID NOS: 34460  
; SOFTWARE: Fast-Seq for Windows Version 4.0  
; SEQ ID NO 272  
; LENGTH: 3841  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-124-367A-272  
Query Match 1.9%; Score 150; DB 7; Length 3841;  
Best Local Similarity 17.5%; Pred. No. 0.03;  
Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;  
Qy 220 HLOATAQRYLHVYVGNRTASSETLPGLDGSAPSQADMLFYLVDFDYGRSNLKTTPAPFST 279  
Db 1979 HLKFTQDVYSAVV-KENSTEATLAVITAGNIPNEPLFYHILNPDRRFKISRTSGVLST 2037  
Qy 280 TGSWLCHQDRFSRYEYGFETRRLCRQVLMYHQLQALD--SKITEHNGPTLVSRLLNY 337  
Db 2038 TGTPFDREQQ-----EAFDVVVEVTEHKKPSAVAHVV--- 2069  
Qy 338 DESAIASTLVFVRVHGHEQDGNV---VTLPLELAYQD----- 372  
Db 2070 -----VKVIEDQNDNAPFVNLPYAVVYVDTEVGHVIRYVTAVDRSGRNGE 2118  
Qy 373 ---FSPRRHAWQPMVDLANFNAIQRMQLVDLKGEGPLGLLYQDKGAWWYRSQRLGEIG 429  
Db 2119 VHYLYKEHHEHFQ-IGPLGEISLKKQFELDTLNKEYLTVTVVAKDG-----G 2163  
Qy 430 SDAVTWKMQPLSVI-----PSLOSASLVLDINGDQGLDWMVITGPCLR 472  
Db 2164 NPAFSAEVIPIITVMNKAMPFPEKPFYSABIAESIQVHSPVHVQAN-----SPEGLK 2216  
Qy 473 GYHSQRPDGSWTRFT-----PLNAL-PVEY-THPRAQLA-----DLMGA----- 509  
Db 2217 VFISITDGPSPQTFINFTNCTVINVIAPLDFEAPKLSIRATDSLTAHAHAEVFDIIV 2276  
Qy 510 -----GLSDVLIGPKSVRLYANTRD----- 530  
Db 2277 DDINDNPPVFAQQSYAVTLSEASVIGTSVQVVRATSDSPNRGSIQYMGFNHSHKSHDHF 2336  
Qy 531 -----GPAKGQVQSGDITLPVPGADPRKLVAFSDVLSGQAHLEVSATKVTWCW 581  
Db 2337 HVDSSSTGLISLRLTLDYEQSRQHTIFVRAVDGGMPTLSSDV-----IVTVDTDLNDN 2389  
Qy 582 PNLGRGRFGQPIITLPGFSQAPATE---FNPAQVYLAD-----LDG----- 617  
Db 2390 PPL-----FEQOIYEARISEHAPGHGFTVCYKAYDADSSDIDKLOYSTLSGNDHKGHFVIDS 2445  
Qy 618 -SGPTDLIYVHTNRDLDFLNKSGNGFAPVTLRPEGLRFDHTCQLOMADVQGLGVASLI 676



Db 2446 ATGIIITLNLHRHAKPYS-----LNLVSVDGV-FRSSTQVHVTVIGG-----N 2489

Qy 677 LSVPHMSPHHRCDI/TNNKP-----WLNEMNNMG-----VHTLYRSSQFWLDEKA 726

Db 2490 LHSAPFLQNEVEVELAENAPLHLLVMEVKTITDGDGSIYGHVTHYI/VNDFAKDRFVINEBG 2549

Qy 727 AALT-----GQTPVCVLPFPPIITLWOTETEDISGNKLVTT 763

Db 2550 QIFTLKLDRETPEAKVSVRLMAKDAAGKVAFC-----TVNVILTDDNDNAPQFRAT 2602

Qy 764 LRY-----ARG-----AWDGRERFRGFVVEQTDHQLAAGNAPERTPPALTKN 808

Db 2603 -KYEWIGSSAAKGTISVVKVLASDADEGSNADITVAIEADSESVKENLEINKLSGVIITK 2661

Qy 809 WYATGLP-----VIDNA-----LSTYWRDDQAFAGSPRFTTWQDNKVPL 850

Db 2662 ESLIGLENEFFTFVRAVDNGSPKESVVLVVKILPPQMQLPKFSEBFTFTVSEDVPI 2721

Qy 851 TPEDDNSYFNRAKGLQLRLSELGLDDSTNKHVPYTVTFERSQV-----RRLQHTDSRY 907

Db 2722 GTEIDLIR-----AHSHTGTVLSLVKGNTPESNRDESVIDRQSGRLKLEKSLDHETTKW- 2776

Qy 908 VLMSVVSZSNVHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPRQQAINLYP--- 964

Db 2777 -----YQFSILARCTODDHEMVASVD-----VSIQVANDNSPVFESSPYEA 2819

Qy 965 ---DTLPDKLLANSYDDQORQLRLTYQQSSWHHLTNTNVRVLGLPDSRSDIFTYGAENV 1021

Db 2820 FIVENLFG-----GSRVIQIRASDADSG-----TNGQV-MYSLDOSQSVESFALNM 2867

Qy 1022 PAGGLNL---ELLSDKN-----SLIADKPREVLGQOKTAYTDGQNTPLQPTTQALIAF 1074

Db 2868 ETGMITTLKELDHKRDYQIKWASDH-GEKILQSSTAI/VBVTVDVNDSPPR----- 2920

Qy 1075 TETTVFNQSTLSAFNGSTPSDK-----LSTT---LEQAGYQCTNVLFRTEGDKVWV 1123

Db 2921 -----FTABIKYGTVSDEDDPOGGVIAILLSTTDADSEENRQVTFI---TGGDPL-- 2967

Qy 1124 AHHGYTDYGTAAQFWR-----POKOSNTQLTGKITLIWDANYCVVVQTRDAAGLTTS 1175

Db 2968 ---GQPAVETITQNEKVVVKFPLDKRDNVILT---IT-----ATD 3003

Qy 1176 AKYDWRFLTPVLQTDINDNOHL-----ITLDALGRPITLRFWGT-----NGKMTG 1221

Db 3004 GTFSSKAIVEKVLNDANDNSPVCKTLYSDTIPEDVLPCKLIMOISATDADIRSNABEIT- 3062

Qy 1222 Y-----SSPEKASFSPSDVNAIEELKPLPVACQVYAPESWMPVLSOKTFNRLAEQDWQ 1277

Db 3063 YTLGSGAEKFLNPDGT-----ELKTSTPL-----DREEQ 3093

Qy 1278 KLYNARIITPD--GRICTLAYRRWQSQKAIPLISILN-NGPRL--PPHSL----- 1324

Db 3094 AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPESADPYAITVFNTEP 3143

Qy 1325 -TLTTRDYDHPPEQIROQVVS-----DG-----FGRLLQAA-----A 1357

Db 3144 GTLLTRVQATDADAGLNKILYSLIDSADGQFSINELSGIILEKPLDRELQAVVTLK 3203

Qy 1358 RHEAGMARQNEGSLIINVQHTENRWATGTEYDNKGQPIRTYQY 1405

Db 3204 AVDQGLPRRLTATGTVIIVSVLDI-----NDNPPVFYREY 3238

RESULT 7

US-11-113-424-53

; Sequence 53, Application US/11113424

; Publication No. US20050260713A1

; GENERAL INFORMATION:

; APPLICANT: Gangolli et al.

; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same

; FILE REFERENCE: 21402-225

; CURRENT APPLICATION NUMBER: US/11/113,424

; CURRENT FILING DATE: 2005-04-21

; PRIOR APPLICATION NUMBER: 60/256,704

; PRIOR FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 60/311,590

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: 60/257,314

; PRIOR FILING DATE: 2000-12-20

; PRIOR APPLICATION NUMBER: 60/311,613

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: 60/315,617

; PRIOR FILING DATE: 2001-08-29

; PRIOR APPLICATION NUMBER: 60/307,506

; PRIOR FILING DATE: 2001-07-24

; PRIOR APPLICATION NUMBER: 60/322,358

; PRIOR FILING DATE: 2001-09-14

; PRIOR APPLICATION NUMBER: 60/294,075

; PRIOR FILING DATE: 2001-05-29

; PRIOR APPLICATION NUMBER: 60/288,153

; PRIOR FILING DATE: 2001-05-02

; NUMBER OF SEQ ID NOS: 190

; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 53

; LENGTH: 2515

; TYPE: PRT

; ORGANISM: Drosophila melanogaster

US-11-113-424-53

Query Match 1.8%; Score 139; DB 7; Length 2515;

Best Local Similarity 16.5%; Pred. No. 0.11;

Matches 236; Conservative 161; Mismatches 446; Indels 586; Gaps 59;

Qy 1 MONSDPSITELSLPKGGGATGMEALTTPGP-----DCMAALSPLPISAGRGVAPAPT 56

Db 1027 VRDNDYSQPELNWE-----AVVSGERCLPDEAHCGDGLAKDAKLAYPEKGIATSSDN 1082

Qy 57 LNVNSGA-----GNSPFGLGW-----DCNVMTIRRRTHFGVPHYDE 92

Db 1083 LYFADGTNIRMVDRDDGIVSTLIQNMHMKSHWKPIPCBSTLKEHMLRWPTLAVSPMDN 1142

Qy 93 TDTFLG-----PEGEVLVADQPRDESTLOGINLGATFTVTGYRSLRSHFSRLEW 144

Db 1143 TLHIIDDHMLRMTDPGRVVRVISGRP-----LHCATASTAYDTDLATHATLV--- 1189

Qy 145 QKTTGTGTDWLIYSPDQVHLLGKSPQARISNPSQTTQTQWLLLEASVSRGEGEYIYQY 204

Db 1190 MPQS-----IAFGPLGELY-VAESDSQRINRVVIGTGRDIAPFAGAEKSCNCL---- 1237

Qy 205 RAEDDTGCEADEITHLQATAQ-----RYLHIVYVGNRTASETLPGDGSAPS 252

Db 1238 -----ERGDCDCEABEHLATSAKFNTIAALAVTPDSHVHIAOQANYRIRSVMSIPEASPS 1293

Qy 253 -----QADMLF-----YLVFDYGRSNNLKTTPPAFTTGTGSLC 285

Db 1294 REYEIYAPDMQEIYIFNRFQGHVSTRNLTGTTVFTYVNTVNTSGKLSLTVTDAAGNKVF 1353

Qy 286 RQDRFSRYEYGFETRRLCLR-----QVLMYHHLQALDS-KIT-EHNGTILVSR----- 332

Db 1354 LLRDYTSQVNSIENTKGOKCLRMTRMKMLHLSLTPDNPVNTYEHGPTGLLRTKLDSTG 1413

Qy 333 --LILNLYDE-----SASTAST-----LVFVVR- 350

Db 1414 RSYVNYNDEFGRLTSAVTPTGRTVRIELSPDLSVKGQVKNVAKEMSLLTQCATVIRN 1473

Qy 351 -----RVGHEQDGNVVTLLPP-----LELAYQDFSPRHHAHWQPMQMDVLANFNAIORQLV 399

Db 1474 GAESRITVDMGSGTTSITPWGHNLOMEVA-----PYTILA-----EGS 1512

Qy 400 DLKGEGLPGLLYQ-----DKGAWWY--RSAQRL-----GEIGSDA--- 432

Db 1513 PLLGESYVPKQRTETAGDLANRFEMRYFVRROQPLQAGKSKGPPRPVTEVGRKLRVN 1572

Qy 433 -----VTWEK-MQPLSVIP----- 445

Db 1573 GDNVLTLEYDRETSQSVVMVDDKQELNVYDRTSRPISFRPQSGDYAYVDLEYDRFRL 1632

QY 446 -----SLQSNASLVNDIN-GP-----GOLDWVITGPGGLRGVHSORPD 480  
Db 1633 VSWKGVLEAYSPDRNGRLNEIKYGDSGTWVYAFKDMFGSLPKVTPTRRSYLLQYDD 1692  
QY 481 G-----SWTRFTPLNALPVE-----YTHPRAQ 502  
Db 1693 AGALQSLTTPRGHTHAFSLQTSGLGFFKYYQYSPINRHPFELLYNDEQOILLAKIHPHOSGK 1752  
QY 503 LADLMG-AGLSDLVLIGPKSVRLVANTRDGPAGKGVVQSGDITLPPGADPRKLVAF-- 559  
Db 1753 VAFVHDTAGRLTETLAGLSSTHYTYQDTTSLVKSVEQVE-----PGFELREFKHYA 1804  
QY 560 -----SDVLGS-----GOAHL-----VEVSATKVTCPNLRGRGRFG 590  
Db 1805 GILDEKLRFGSKNSLASARYKAYDGNARLSGEMAIIDDKELPPTTRYKYSONJQLLEV 1864  
QY 591 Q--PITLPGFSQAPATEFNPQVY--LADLDGSGPTDLIVVHTNRDLIF---LNKSGNGFAE 644  
Db 1865 QDLKITRANFNRTVIQDSAKQFFAIVDYDQGRVKSVMNVKNIDVFRLELDYDLNRRIK 1924  
QY 645 PVTLRPFEGURFD---HTCQLQMAADVQGLGVASLILSVPHMSPHHWRCDLTNNKFWLLNE 701  
Db 1925 SQKTTFGSTAFDKINTYNADGHVVEVLG-----TNNWKYLFDE 1962  
QY 702 MNNMGVHHTLRYRSSQFMDLDEKAAALTTGQTPVCVLPPIHTLMQTEDEISGNKLV 761  
Db 1963 NGNTVG-----VDQGEKFNLYGDI-----GGRVI 1987  
QY 762 TLLYARGAMDGR-----EREPFGFYVEQTDSHQLAQNAPERTPPALTKWYATGL 814  
Db 1988 KVGDFENNYDARGFVVRGEOKYR-----YNNRGQLIHSFERER-----FQSM----- 2031  
QY 815 PVIDNALSTEWDRDDQAFAGESPFTWQDNKQVLPDTPEDDNSRYFNRAKLGQLLRSEL 874  
Db 2032 -----YYDDRS-----RLVAHDNK-----GNTQYY----- 2055  
QY 875 YGLDDSTNKHYVYVT-----EFESQVRRLOHTDSRYPLVLMSSVVERNY 919  
Db 2056 -----ANRTPHLVTHVHPKISRTMKLFYDDRDLIALEHEQRYV----- 2098  
QY 920 HYERIASDPQCSQNTLSSDRFGQPL-----KQLSVQYPRRQOPAINLYPDTLPDKLLANS 975  
Db 2099 -----ATDQNGSLAFFDQNGSIVKEMKRTPGRIIKOTKPEFFVPID 2141  
QY 976 YD-----DQORQLRITYQSSWHHL-----TNNTVRVGLPDSTRSDIFTYGAEVNPAGGLNL 1028  
Db 2142 FHGGLIDPHTKLXYTEQRYDPHYGQWMTPLWETLATEMSHTDVFYRYHN----- 2193  
QY 1029 ELLSDKNSLIADDPREVL-----GQQKTAVTDGQNTTP 1062  
Db 2194 -----NDPINPNRPNQNTMIDLSWQLFGYDLNMMQSSRYTKLQAQYTP 2236

## RESULT 8

US-11-113-424-14  
; Sequence 14, Application US/11113424  
; Publication No. US20050260713A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli et al.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-225  
; CURRENT APPLICATION NUMBER: US/11/113,424  
; CURRENT FILING DATE: 2005-04-21  
; PRIOR APPLICATION NUMBER: 60/256,704  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/311,590  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/257,314  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/311,613  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/315,617

PRIOR FILING DATE: 2001-08-29  
PRIOR APPLICATION NUMBER: 60/307,506  
PRIOR FILING DATE: 2001-07-24  
PRIOR APPLICATION NUMBER: 60/322,358  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/294,075  
PRIOR FILING DATE: 2001-05-29  
PRIOR APPLICATION NUMBER: 60/288,153  
PRIOR FILING DATE: 2001-05-02  
NUMBER OF SEQ ID NOS: 190  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 14  
LENGTH: 2769  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-11-113-424-14

Query Match 1.7%; Score 136; DB 7; Length 2769;  
Best Local Similarity 18.2%; Pred. No. 0.21;  
Matches 332; Conservative 211; Mismatches 563; Indels 714; Gaps 101;

QY 66 SPFGLGHDCNMVIRRTTHFGVPH-----YDETDTFLGPEGEVLVVADQP----- 110  
Db 866 NPLCLG-SPNPLDIIQETQVPSQQNLHSPYDRIKFLVGRDSTHIIIPGENPFDGHCACVI 924  
QY 111 -----RDESTLOGINLG-ATFTVTGYR-SRLESFHS-----RLEYWQPKTKG 151  
Db 925 RGQVMTSDGFLVGVNLSFVNPLFGYTI SRDGSFLVTNGGISIILRER-APFITQE 983  
QY 152 TDFWL-----IYSPDQVHLLGKSPQARISN-----PSQTTQAO----- 186  
Db 984 HTLWLPWDRFPVMEIIMRHEENEIPSCDLSNFARPNPVSPSLTSFASSCAEKGPVP 1043  
QY 187 --WLEASVSRSRGEIYYQRAEDDTCEA---DEITH-----HLQATQRYLHI 231  
Db 1044 EIQAQOEIISGCKMFLSYLSRTPGYKSVLRISLTHPTTIPFLMKVHLMVAVEGRLEFR 1103  
QY 232 VYGNRTASETLPGLDGSAAPSQADWLFYVFD---YGERSNLKTTPAFSTTGSWLCRQ 287  
Db 1104 KWF-----AAAP---DLSYFFWTKDVTYVQKVFGLS---EAFVSVG----- 1139  
QY 288 DRFSRYEY-----GFEIRTRLCROVLMYHHLQALDSKITEHNG----- 326  
Db 1140 -----YBESCPDLILWEKRTTVLQGYEIDASKLGWSLQKHALNIQSGIL-HKNGEN 1193  
QY 327 -----PTLVSRLLN-----YDESALASTLVFVR 351  
Db 1194 QFVSQQPFVIGSINGNGRRRSISCPSCNGLADGNKLLAPVALTCGSDGSLYVGDFTVIRR 1253  
QY 352 VGHEQDGNVVTLPPELAYQDFGPRH---HAHWQPMVLANFNAIORWQLVDLKGELPG 408  
Db 1254 I--FPSGNVTNI--LELRNKDFRHSHPAKYIYATDPM----- 1289  
QY 409 LLYQDKGAWYRSAQRIGEIGSDAVTWKMQPIASVIPSLOSNASLVLDINGQQLDWITG 468  
Db 1290 -----GAVFLSDNSRRVF--KIKSTVVVKDLVKNSEVWAGTQDCLPFDTR 1335  
QY 469 PGLRGYHSQRPDGSWTFPTLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANT 528  
Db 1336 CG-----DGG-----KATEATLTNPRGTITVDKFGI-----IY--- 1362  
QY 529 RDGFAGK---KDVVQSGDITLVPFGADPRKLVAFSDVLGSCQAHLEVSATKVTCPNLRG 585  
Db 1363 ---FVDGTMRIRDQNGII-----STLLGSNDL---TSARPLSC----- 1395  
QY 586 RGRFGQPIITLPGFSQAPATEFNPQVYLAIDLGGSGFTDLIYYVHTNRDLIFLNKSGNGFAEP 645  
Db 1396 -----DSVMDISQVLEW-PTDLAINPMDNS-----LYVLDNN----- 1427  
QY 646 VTLRPFEGLEFDTQCLOQADVQGLGVASLILSVPHMSPHHWRCDLTNNKFWLLNEMNN 705  
Db 1428 VWLQISE---NH---QVRIVAG-----RPMH---CQVPGIDHFLUSK----- 1460





```
QY 1010 RSDIFTY---GAENVVAG-----GLNLELLS-----1032
Db 1724 NPNVGTIVVSKSKONTGTLFANFGNNQNVGGTVFYNQGVDCADCGMYRRSMQYFGIP 1783
QY 1033 -----DKNSLI-----ADKPXYLQOQKTAFTDQNTTPL 1063
Db 1784 VNESDPPYDHVDGNATVQWVFPNGDKWRPAPYAPDTKLQFKGYQITNDVQAO-----1838
QY 1064 QTPTR-----QALIAFTETTVFNQSTL-----SAFNGSIPSDKSLTTLQAGYQO 1108
Db 1839 --PTGVYSFGKGLCVCDAPLNLTRTSGVNYSGANLIGNSYTGAI--DIKQGIIVFPPEVEQ 1894
QY 1109 TNYLFPRTGCEKWAHGHYTDYG--TAAQFWRPQKSNTO--LTGKITLIWDANYCVVQO 1165
Db 1895 TVYLF--NTGTRDQWRKLGSTVSGYRAGQYLSVPKNTAGQDNLPRI-----PSMHSFLVK 1949
QY 1166 TRDAAGLTTSAKYDWRFLTPVOLTDINDNQHLITLDALGRPTTLRFMTGTEGKMTGYSSP 1225
Db 1950 MONGASCTLIQIYD-----KLLKNTTVNNG-----NGTQITWRSQ 1984
QY 1226 EKASFPSPSDVNAALIELKPLVPAQCQVYAPESMMPVLQKTFNRLAQDWOKLYNARI 1285
Db 1985 NSGSANMPSLV-----MDVLGNESADRL-----W-----IF 2010
QY 1286 TE-----DGRICITLAYRRVWQSKAIPOLISLLNNG-----PRLPPHSIT 1325
Db 2011 TDGGLSGFGDNGWGRKLT-----EKLGSQLYANMSDIGNDKFQVAGVPEL--NNLL 2059
QY 1326 LTTDRYDHDPEQIRQOVVFSDFG---GRLLQAAARHEAGMARQNEGSLIINVQHTEN 1382
Db 2060 I---GFDADKQGYTLFALSDFHFAKGVFLEDLSR---GVTRRVVDCGSYSFADKRGDS 2113
QY 1383 RMAVGTGRTEYDNK 1395
Db 2114 --GARFRLSYDEE 2124

RESULT 11
US-10-453-372-718
; Sequence 718, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See 'File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 718
; LENGTH: 2591
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-10-453-372-718

Query Match      1.6%; Score 127; DB 6; Length 2591;
Best Local Similarity 19.1%; Pred. No. 0.94;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

QY 91 DETTFLGPEGEV-----LVVADQPRDESTL-----QGINLGATTVTGYRSLRSH 137
Db 309 DSSAFISPOGFMAFPFGLTTLNMTDQSGNEANMVCISIQKPSRTSFIAPTEENDYIVLNTS 368
QY 138 FS-----RLEV-----WQPKTTGKTDFWLYSPDGQV-----HLGKSPQARISNPSQTT 182
Db 369 FSTFLVCNIDIGHIQPWQO-----ILALYSOSPILERSHLLSETFQL-----411
QY 183 QTAQWLLASVSSRGEIYYQYRAB---DDTGCEADEITHHLOATAQRYLHIVVYGNRTA 239
Db 412 ---YKYKQVAPKPEDIFTNIEADLRADPSLWMDQDISLQNLRTATTFTSLTIQYSSDA 467
QY 240 SETLPGLDGSAPSQADWLFYLVFDYGERSNLIK---TPPAFSTTG-----SW 283
Db 468 QITLPRAE--MRPVRKHWTMI-----SRDNTKLEHTVLVGGTVGLNCPGQGDPTPHVDM 520
QY 284 LCRQDRFSRYEGFE-----IRTRRLCRQVL-----MYHHLQALDSDKITEHNGPTLVS 331
Db 521 LLADGSKVRAPYVSEDGRILLIDKSKLEQWADSFDTGTYHCIS-----565
QY 332 RLILNYDESATIASTLVFVRRVHGEGDQGNVVTLPPELAYQDFSPRRHAHW--QPMDVLAN 389
Db 566 ---NYDDADILTY-----RITVVEPLVEAYQE--NGIHHTVFIGETLDPCH 607
QY 390 FNAIQ---RQWLDLKGEGLPG---LLYQDKGAWWYRQAQLGEIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILOV 646
QY 444 IPSLQSNASLVINDGQDQWVITGPGLRGVHSORP---DGSWTRFTPLNALPVEYTHPR 500
Db 647 TPKDQGYRCAVAPPSG--VDFLIIFQVSVK--MKGQRPLEHGE-----TESGLDESPI 698
QY 501 AQLADLMGAGLSDLVLI-----AKGVVQSGDITLVPFGADPRKLVAFSDVLG-----SGQAHVLE 572
Db 759 ARRIPDQHWAAALLEKAKGNAMPDKRENTVSP---PPVVTQLPNIPGEBEDSSGMLALHE 815
QY 573 ---VSATKVTCWPNL-----GRGRFGQPIITLPGFSQAPATEFNPAOVYLADLDGSGPTDL- 623
Db 816 EFWVPATKALNLPARTVTADSRITSDSPMTNINY---GTEFSPV--VNSQILPPEEPTDFK 871
QY 624 --IYVHTNRLDIFLNKSGNGFAEPVTLR-----FP-----EGLRFDH-----658
Db 872 LSTAIKTTAMSKNINPTMSSQIQGTTNQHSSVTVPFLLLGATEFQDSQDMGRHFQSRP 931
QY 659 --TCOLOWADVQGLGVAS-----LILSVPHMSPHHWRCDLTNMKFWLLNMMNNMGVHHT 711
Db 932 PITVTRTIKDVNVNMLSSSTTNKLLLESVNTNHSQO-----TSVRE--VSEPRHNHYGHT 984
QY 712 LRYBSSQFWLDEKAAALTTCQTVCYLPFPFIHTLWQTEDEISGNKLVTLTLRVARGAW 771
Db 985 TQILSTSTFPSPDPHTAA-----HSQFP-----IPRNTVINPLFR---1020
QY 772 DGRERERFGYVQGTDSHQLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWRDQA 831
Db 1021 FGRQKIGRGRI-----ISPVRT--PVLRRHYSI-----FRSTTRCSSEKS 1061
QY 832 FAGFS-----PRFTWQ-----DNKDVPLTDEDDNSRTWFNRA 864
Db 1062 TTAFSATVLNVTCLSCLPRRLTATAALSFPSSAAPITFPKADIARVSEEST-----1114
QY 865 LKGQLLRSELGLDD--STNKHVPVTVTFFRSQVRRLOHTDSRYFVLWSSVVESR---N 918
Db 918
```

Db 1115 ---TLVQNPLLLLENKPSVEKTTT-TIKYFRTEISQVTPGTGAVMYTAFTSIPMEKTHKVN 1170  
Qy 919 YHYERIASDPQCSQNIITLSSDRFG---QPLKQLSV-QYPRQOPAINLYDTPDKLLA 973  
Db 1171 ASYPRVSTNEAKDSVITSSLSGAIKTPPMIIITAITFRSRRKIPW-----QQNFV 1221  
Qy 974 NSYDDQOR---QLRLTYQQSS-----MHHLTNNTVRVLGLP-----1006  
Db 1222 NNHNPGLRNLQHKVSLQKSTAVMLPKTSPALPQORQSLPSHHTTKTHNPSLSLTKKELP 1281  
Qy 1007 -----DSTRSDIFTYGA-----ENVAGGLNLELLSDKNSLIADDKPREY 1046  
Db 1282 PPLNPMPLPSIISKDSSTKSIISTQTAIPATPTTPPASVITYETQTERSRAQTIQREOE- 1340  
Qy 1047 LGOQKATYDQONTPLQ-----TPTROA---LIAFTETTVNQSTLSAFNGSIPSDK--L 1097  
Db 1341 --POKKNRTD-NTISPDQSSGFTTPTAMTPVLTAAETSV--KPSVSAPFTHSPENTTGI 1395  
Qy 1098 STTL 1101  
Db 1396 SSTI 1399  
RESULT 12  
US-10-453-372-716  
; Sequence 716, Application US/10453372  
; Publication No. US2006000323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: Curaseq version 0.1  
; SEQ ID NO 716  
; LENGTH: 2602  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-716  
Query Match 1.6%; Score 127; DB 6; Length 2602;  
Best Local Similarity 19.1%; Pred. No. 0.95;  
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;  
Qy 91 DETDTFLQPEGEV-----LWADQPRDESTL-----QGINLGATFTVTVGYSRLSEH 137  
Db 309 DSSSAFTSPQGFAPFGSLTNMTDQSGNEANVCSIQKPSRTSFIATFENDYIVLNTS 368  
Qy 138 FS-----RLEY-----WQPKTKGTFWLYSPDQV-----HILGKSPQARIENPSGTT 182  
Db 369 FSTFLVCNIDYGHIQPVNQ-----ILALYSDSPILERSHLLSETPQL-----411

Qy 183 QTAOWLLEASVSSRGEQIYYQYRAE---DDTGCADBIITHLOATAQRYLHIVYGNRTA 239  
Db 412 ---YKYKQVAPKPEDIFTNIEADLRADPSWLMDQOISLQLNRTATFTFLQIQYSSDA 467  
Qy 240 SETLPGLDGSAPSQADWLFYLVFDYGBRSNNLK---TPPAFSTTG-----SW 283  
Db 468 QITLPRAB-MRPVHKWTMI-----SRDNNTKLEHTVLVGGTVGLNCPGQGDPTPHVDM 520  
Qy 284 LCRODRFSRYEYGEF-----IRTRRLCRQVL-----MYHHLQALDSKITEHNGPTLVS 331  
Db 521 LLADGSKVRAPYVSEDEGRILIDKSGKLEQWADSFDTGVVHCITSS-----565  
Qy 332 RLILNYDESATASLTAVFVRVGHEDGNVVTLPLELAYODFSRHHAAHW---QPMVDVLAN 389  
Db 566 ---NYDDADILTY-----RIIVVELEVEAYQE-NGIHHTVFIGETILDLPCH 607  
Qy 390 FNAIQ-----RWQLVDLKGEGPLG---LLYQDKGAWWYRSQAQLGIEIGSDAVTWKMQPLSV 443  
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRLIQV 646  
Qy 444 IPSLOSNASLVDINGDQLDWITPGRLRGVHSORP---DGSWTRFTPLNALPVEYTHPR 500  
Db 647 TPKDQGYRYCVAANPSG-VDFLIFQVSVK-MKGORPLEHDGE-----TEGSGLDESNPI 698  
Qy 501 AOLADLMGAGLSDLVLI-----GPKSVRLYANTRDGF---532  
Db 699 AHLKEPPQAQLRTSALMBAEVGKTSSTKXHNRYRELTQRRGDSHRRFRHRPPS 758  
Qy 533 -----AKGDVVQSGDITLPPVGADPRKLVAFSDVLG-----SQOAHLVE 572  
Db 759 ARRDPQHWAAALLEKAKKNAMPDKRENTVSP---PPVVTQLPNIPEGEDSSGMLALHE 815  
Qy 573 ---VSAIKVTCWPL-----GRGFGQIPITLPGSQPATEENPAQVVLADLDGSGPTDL- 623  
Db 816 EFMVPATKALNLPARTVTDASRTISDSPMTNINY---GTEFSPV-VNSQLLPPEPTDFK 871  
Qy 624 --IYVHTNRLDIFLNKSGNGFAEPVTLR-----PP-----EGLRFDPH-----658  
Db 872 LSTAIKTAMSKNINPTMSSQIQGTTNQHSTVFPFLLLGATEFQDSQDMGREHFQSRP 931  
Qy 659 --TCOLOMADYQGLGVAS-----LILSVPHMSPHHRCDLTNMKPMLNENNNMNGVHHT 711  
Db 932 PITVTRMIKDVNVKMLSSTTNKLLESVNTNSHQ-----TSVRE--VSEPRHNHFYSHT 984  
Qy 712 LRYSSSQFWLDEKAAALITGQTPCVLPPPIHTLMQTEDEDEISGNKLYTTLAYARGAW 771  
Db 985 TQILSTSTFPSPDPTAA-----HSQFP-----IPRNSTVNIPLFR---1020  
Qy 772 DGREREPFGYVEQTDSHQLAQGNAPERTPPALTKNMYATGLPVIDNALSTEYWRDDQA 831  
Db 1021 FGORKIGGRRI-----ISPYRT-PVLRHRYSI-----FRSTTGSSEKS 1061  
Qy 832 FAGFS-----PRFTTWQ-----DNKQVPLTPEDDNSRYWFNRA 864  
Db 1062 TTAFSATVLNVTCLSCLPRERLTATAALSPSPAAPITFPKADIARVPSEST-----1114  
Qy 865 LKGQLLRSELYGLDD---STNKHVPYTYTEFRSQVRRLQHTDSRYPLWSSVSES---N 918  
Db 1115 ---TLVQNPLLLLENKPSVEKTTT-TIKYFRTEISQVTPGTGAVMYTAFTSIPMEKTHKVN 1170  
Qy 919 YHYERIASDPQCSQNIITLSSDRFG---QPLKQLSV-QYPRQOPAINLYDTPDKLLA 973  
Db 1171 ASYPRVSTNEAKDSVITSSLSGAIKTPPMIIITAITFRSRRKIPW-----QQNFV 1221  
Qy 974 NSYDDQOR---QLRLTYQQSS-----MHHLTNNTVRVLGLP-----1006  
Db 1222 NNHNPGLRNLQHKVSLQKSTAVMLPKTSPALPQORQSLPSHHTTKTHNPSLSLTKKELP 1281  
Qy 1007 -----DSTRSDIFTYGA-----ENVAGGLNLELLSDKNSLIADDKPREY 1046  
Db 1282 PPLNPMPLPSIISKDSSTKSIISTQTAIPATPTTPPASVITYETQTERSRAQTIQREOE- 1340

```

QY 1047 LQOQKATYDQNTTLPQ-----TPROA--LIAFTETTVFNQSTLSAFNGSIPSDK--L 1097
Db 1341 --PQKNRTD--PNISPDQSSGTTTPTAMTPPVLTAAETSV--KPSVSAFTSHSPPTTGTI 1395
QY 1098 STTL 1101
Db 1396 SSTI 1399

RESULT 13
US-10-453-372-666
; Sequence 666, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 666
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-666

Query Match 1.6%; Score 127; DB 6; Length 2617;
Best Local Similarity 19.1%; Pred. No. 0.96;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

QY 91 DETDFLPGEGV-----LVVAQPRDESTL-----QGINLGAFTVGYRRLSH 137
Db 309 DSSAFISPGFMFPGSLTNMTQSGNEANMVCSIQKPSRTSPIAFTENDYIVLNTS 368
QY 138 ES-----RLEY-----WPKTKTGDFTWLIYSPDGV-----HLGKSPQARISNPSQT 182
Db 369 FSTFLVCNIDYGHIFVWQ-----ILALYSDSPILERSHLLSETPOL----- 411
QY 183 QTAOWLLEASVSRGEQIYYQVRAE---DDTCEADEITHLQATAQRYLHIVVYGNRTA 239
Db 412 ----YKYKQVAPKPEDIFTNEADLRADPSWLMQDQISLQNRATTFTSTLIQYSSDA 467
QY 240 SETPLGLGSAQSADWLFYLVFDYGRSNNLK-----TPAPFTTG-----SW 283
Db 468 QITLPRAE--MRPVKHKWTMI-----SRDNNTKLEHTVLVGTVGLNCPGQGDPTPHVDW 520
QY 284 LCRQDRFSRYEYGF-----IRTRLCQVL-----MYHHLQALDLSKITEHNGPTLVS 331
Db 521 LIADGSKVRAPVSEDRILIDQKLEQMAADSPDTGVYHICSS----- 565
QY 332 RLILNYESATASTLVFVRRVGHEQDGNVVTLPPLLELAYQDFSPRRHAW--QPMVDVLAN 389

```

```

Db 566 ----NYDDADILTY-----RITVVEPLVEAYOB--NGIHHTVTFIGETLDDLCH 607
QY 390 FNAIQ-----RQVLVDLKGEGLPG--LLYQDKGAWMYSAQRLGEIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILOV 646
QY 444 IPSLQSNASLYVDINGDQLDWITGPGRLGYHSQRP---DGSWTRFTPLNALPVYTHPR 500
Db 647 TPKDQGYVRCVAANPSG--VDELIQVSVK--MKGQRPLEHDE-----TEGSLDESNP 698
QY 501 AOLADLMAGLSDLVLI-----GPKSVPLYANTYRDFG--- 532
Db 699 AHLKEPPGAQLRTSALMEAEVKGHTSSTSKHNYRELTLQRRGDSHRRFRNRHFPSP 758
QY 533 -----AKGKDVVQSGDITLPPVGADPRKLVAFSDVLG-----SGOHLVE 572
Db 759 ARRIDPQHWAAALLEKAKKNAMPDKRENTTVSP---PPVVTQLPNIPGEDSSGMLALHE 815
QY 573 ---VSATKVTCWPNL-----QGRFGQPIILPGFSQPATEFNPAQVYLADLDGSGPTDL- 623
Db 816 EFMVPATKALNLPARTVTADSKTISDSPTMINY---GTEFSPV--VNSQILLPPEPTDFK 871
QY 624 --IVHTNRLDIFLKNSGNGFAEPVTLR-----PP-----EGLRFDH----- 658
Db 872 LSTAIKTTAMSKNINPTMSSQIQGTTNQHSSTVPFLLLGATEFQSDQMGREHFQSRP 931
QY 659 --TCOLOMADVOGLGVAS-----LILSVPHMSPHHWRCDLTNMMKFWLLNMMNNMGVHT 711
Db 932 PITVRTMIKDNNVWMLSTTNKLLSVNTNSHQ-----TSVRE--VSEPRHNFYSHT 984
QY 712 LRYRSSQFWLDEKAAALTTQQTVPVYLPFPIHTLMOTETEDEISGNKLVTLRLVARGAW 771
Db 985 TQILLSTSTFPSPDPTAA-----HSQFP-----IPRNSTVNIPLFR--- 1020
QY 772 DGRREPFPGYVEQTDSHQLAQNAPERITPALTKNWYATCLPVIDNALSTEYWRDOA 831
Db 1021 FGRQKIGGRGRI-----ISPYRT--FVLRHRYSI-----FRSTTRGSSSKS 1061
QY 832 FAGFS-----PRFTTWQ-----DNKDVPLTPEDDNSRYWFNRA 864
Db 1062 TFAFSATVNLVTCLSCLPRELTTATAALSFPSPAIPFPKADIARVSEEST----- 1114
QY 865 LKGQLLRSELYGLDD--STNKHVPYVTFEFSQVRRLOHTDSRYVLSVSSVVER---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTP--TIKYFTEISQVTPGTGAVMTYAPTSPMEKTHKN 1170
QY 919 YHYERIASDPOCSQNTLSSDRFG-----OPLKQLSV--CYPRQOPAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSLSGAITKPPMTIITAITFRSRKIPW-----QQNFV 1221
QY 974 NSYDDQQR---QLRLTYQSS-----WHLLTNNTVRLGLP----- 1006
Db 1222 NNHAPKGLRNQHKVSLQKSTAVMLPKTSPALPQOSLPSSHHTTKTNPGSLPTKXELP 1281
QY 1007 -----DSTRDIFTYGA-----ENVPAGLNLLELSDKNLSIADDPREY 1046
Db 1282 FPPLNPLMPSIISKSDSTKSIISTQTAIPATTPPTFPASVITVETOTERSAQTIQREOB- 1340
QY 1047 LQOQKATYDQNTTLPQ-----TPROA--LIAFTETTVFNQSTLSAFNGSIPSDK--L 1097
Db 1341 --PQKNRTD--PNISPDQSSGTTTPTAMTPPVLTAAETSV--KPSVSAFTSHSPPTTGTI 1395
QY 1098 STTL 1101
Db 1396 SSTI 1399

```

RESULT 14  
 US-10-453-372-732  
 ; Sequence 732, Application US/10453372  
 ; Publication No. US2006000323A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 732
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-732

Query Match          1.6%; Score 127; DB 6; Length 2617;
Best Local Similarity 19.1%; Pred. No. 0.96;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

Qy 91 DETDTFLQPEGEV-----LVVADQPRDESTL-----QGINLGATFTVTGYSRLESH 137
Db 309 DSSSAFISFQGFAPFGSLTNMTDQSGNEANMVCSIQKPSRTSPIAFTENDYIVLNTS 368
Qy 138 FS-----RLEY-----WQPTTKGTDFWLSPDQGV-----HLGKSPQARISNPSQTT 182
Db 369 FSTFLVNCIDTGHQPVWQ-----ILALYSUSPLILERSHLLSETPQL-----411
Qy 183 QTAQWLLSEASVSRGEQIYYQYRAE---DDTGCEADEITHLQATAQRYLHIVYGNRTA 239
Db 412 ----YYKYQVAPKPEDIFTNIEADLRADPSWLMQDQISLQLNRTATTFTSLQIQYSSDA 467
Qy 240 SETLPGLDGSAPSQADWLFLYLVFDYGERSNNLK---TPPAFTWG-----SW 283
Db 468 QITLFPRAE-MRFPVKHKWTMI-----SRDNNTKLEHTVLVGGTIVGLNCPGQGDPTPHVDW 520
Qy 284 LCRQDRFSRYEYGF-----IRTRRLCRQVL-----MYHLQALDSKITEHNGPTLVS 331
Db 521 LLADGSKVRAEYVEDGRILLDKSKLEQLWADSFDTGVYHCIS-----565
Qy 332 RLILNYDESALASTLIVFRRVVGHEQGNVVTLPPLLELAYQDFSPRRHAHW--QPMVDVAN 389
Db 566 ----NYDDADILTY-----RITVVEPLVEAYQE-NGIHHTVFIGETLIDLPC 607
Qy 390 FNAIQ-----RWQLVDLKGEGLPG--LLYQDGKAWWYRSAQLRGEIGSDAVTWKQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILOV 646
Qy 444 IPSLQSNASLVYDINGDGLDGVITTCPLGRGVHSORP---DGSWTRFTPLNALPVEYTHPR 500
Db 647 TPQGGYTRCVAAWPSG-VDFLIIFQVSVK-MKGQPLEHDE-----TEGSGLDENPI 698
Qy 501 AQLADLMGAGLSDLVLI-----GPKSVELYANTRDGP---532
Db 699 AHLKEPPGAQLRTSALMEAEVKGKHTSSTSKRHNRYRELTQLRRGDSRTHRRFRNRHPPPS 758

```

```

Qy 533 -----AKGDVQVSGDITLTPVGADPRKLVAFSDVLG-----SQOAHIVE 572
Db 759 ARRIDPOHWAALLEKAKKNAMPDKRENTVSP---PPVTQLPNIPREGEDSSGMLALHE 815
Qy 573 ---VSATKVTCPNPL-----GRGRFGQPIITLPGFSQPATBFNPAQVYLADLDGSGPTDL- 623
Db 816 EFMVPATKALNLPARTVADSRITSDSPMTNINY---GTFSVPV-VNSQILPPEEPTDFK 871
Qy 624 --IYVHTNRLDIFLNGSGNGFAEPVTLR-----PP-----EGLRFDH----- 658
Db 872 LSTAIKTTAMSKNINPTWSSQIQGTINQHSSTVFPLLLGATEFQDSDQMGCRGHEHFSRP 931
Qy 659 --TCOLOWADVQGLGVAS-----LILSVPHMSPHHWCIDLTNMKPWLILNENNNMGVHHT 711
Db 932 PITVRTWKIDVNVKMLSSSTNNKLLLESVNTNSHQ-----TSVRE--VSEPRHNFYSHT 984
Qy 712 LRVSSSQFWLDEKAAALTTGQTPVCYLPPFIHTLWQTEDETEISGNKLVTLRYARGAW 771
Db 985 TQILSTSTFFSDPHTAA-----HSQFP-----IPRSTVNIPLFRR--- 1020
Qy 772 DGRERERFGYVEQTDSHQLAQNAPERTPPALTKNWYATGLPVIDNALSTEWRRDQA 831
Db 1021 FGRQKIGRGRGI-----ISPYRT-PVLRHRYYSI-----FRSTTRGSSEKS 1061
Qy 832 FAGFS-----PRFTTWQ-----DNKOVPLTPEDDNSRYWFNRA 864
Db 1062 TTAFSATVNLVNTCULPRERLITATAALSPPSAAPTFFPKADIARVPSEST----- 1114
Qy 865 LKGQLRSELYGLDD--STNKHVEYVTEPRSQVRRLQHTDSRYPLVWSSVVESR---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTP-TIKYFRTETISQVTPTGAVMTYAPTSIPMEKTHKVN 1170
Qy 919 YHVERIASDPQCSQNTLSSDRFG-----QPLKQLSV-QYPRRQOPAINLIPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKPKPMIIAITFRSRRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLELTYQSS-----WHHLTNNTVRVLGLP----- 1006
Db 1222 NNNPKRLRNQHKVSLQKSTAVMLPKTSPALPORQSLPSHHTTKTNPGSLTKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGGLNLELLSDKNSLIADDKPREY 1046
Db 1282 FPLPLNPLPSIISKDSSTKSIISTQTALPATTPTFPASVITYETQTSRSRAQTIQREQ- 1340
Qy 1047 LGQOKTAYTGONTPLQ-----TPTRQA--LIAFTETTVNQSTLSAFNGSIIPSDK--L 1097
Db 1341 --PQKQRTD-PNISPQSSGFTTPTAMTPPVLTAETSV--KPSVSAFTSHSPPEPTTGI 1395
Qy 1098 STTL 1101
Db 1396 SSTI 1399

RESULT 15
US-10-453-372-734
; Sequence 734, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446

```



```

; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 734
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-734

Query Match
Best Local Similarity 1.6%; Score 127; DB 6; Length 2617;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

Qy 91 DETDTFLGPEGEV-----LVADQPRDESTL-----QGINLGATFTVTGYSRRLSH 137
Db 309 DSSAFISFQGFMAPFGLTLMTDQSGNEAMVCSIQKPSRTSPIAFTTEENDYIVLNTS 368
Qy 138 FS-----RLEY-----WPKTTGKTDFWLYSPDQV-----HLGKSPQARINPSQTT 182
Db 369 FSTFLVCNIDYGHIOFVWQ-----ILALYSDSPILERSHLLSETPOL----- 411
Qy 183 QTAOWLLBASVSSRGEQIYYQVRAB---DDTGCEADEITHLQATAQRYLHIVYGNRTA 239
Db 412 ----YKQYQAPKEDIFTNEADLRADPSWLMQDQISLQNLRTATFTSTLIQYSSDA 467
Qy 240 SETPLGLDGSAPSQADWLFYLVFDYGRSNNLK---TPPAFSTTG-----SW 283
Db 468 QITLPRAB-MRPVKHKWTMI-----SRDNNTKLEHTVLVGTVGLNCPGQGDPTPHVDW 520
Qy 284 LCRQDRFSRYEGFE-----IRTRLCROVL-----MYHLQALDSKITEHNGPTLVS 331
Db 521 LLDGSKVRAPVVSDEGRILIDKSKLEQWADSPDTGVYHCIS----- 565
Qy 332 RLILNYDESATLSTLVFVRVGHQDGNVVTLPPLLELAYQDFSPRRHAHW--QPMVDVLAN 389
Db 566 ----NYDDADILTY-----RITVVEPLVEAYQB-NGIHHTVFIFETLDPCH 607
Qy 390 FNAIQ-----RWQLVDLKGEGLPG--LLYODKGAMWYRSQRIIGEIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRLIQV 646
Qy 444 IPSLQSNASLVINDGQDGLDWITCPGLRGVHSORP---DGSWTFPTPLNALPVEYTHPR 500
Db 647 TPKDQGYRYCVAANPSG-VDFLIQVSVK-MKGQRPLEHGE-----TEGSGLDSENP 698
Qy 501 AOLADLMGAGLSDLVLI-----GPKSVRLYANTRDGF--- 532
Db 699 AHLKEPPGAQLTSALMEAEVKGKHTSSTSKRNYRELTLQRGDSTHRRFRENRRHFPPS 758
Qy 533 -----AKGQDVQSGDITLVPVGDAPRKLVAFSDVLG-----SQQAHIVE 572
Db 759 ARRDPQHWAAALLEKAKKNAMPDKRENTVSP---PPVVTQLPNIPGEEDDSSGMLALHE 815
Qy 573 ---VSATKVTWPNL-----GGRGFGQITLPGFSQPATFNPAQVYLADLDGSGPTDL- 623
Db 816 EFMVPATKALNALPARTVTADSTISDSPMTNINY---GTEFSPV-VNSQILLPPEEPTDFK 871
Qy 624 --IYVHTNRLDIFLNKSGNGRAEPVTLR-----FP-----EGLRFDH----- 658
Db 872 LSTAIKTTAMSKNIINPTMSSQIQGTNQHSSVFFLLIGATEFQDSQWQGRGHEHFQSRP 931
Qy 659 --TCOLOMADVQGLGVAS-----LILSVPHMSPHHRCOLTNMKPWLNNMNNNGVHHT 711

```

```

Db 932 PITVRTMIKDVNVKMLSSSTTNKLLLESVNTTNSHQ-----TSVRE--VSEPRHHFYGHT 984
Qy 712 LRYRSSQFMLEDEKAAALTTGQTPVCVLPFFPIHTLMQTEDEISGNKLVTLTRYARGAW 771
Db 985 TQILSTSTFFSPDPHTAA-----HSQFP-----IPRSTVNIPLFR--- 1020
Qy 772 DGRERFEGFYVEQTDSSHQAQGNABERTPPALTUKWYATGLVIDNALSTEVWRDQQA 831
Db 1021 FGRQKIGGRGRI-----ISPYRT-PVLRHRHYSI-----FRSTTRGSSEKS 1061
Qy 832 FAGES-----PRFTTWO-----DNKDVLPTPEDDNSRYWPNRA 864
Db 1062 TTAFSATVNLVNTCLUSCPLPRRLTTATAALSPSAPITFPKADIARVSEEST----- 1114
Qy 865 LKGOLLARSELGLDD--STNKHVPYVTFEFSQVRRLOHTDSRYFVLWSSVSES---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTP-TIKYFRTEISQVTPGTGAVMTYAPTSPMEKTHKVN 1170
Qy 919 YHYERIASDPQCSQNIYLSDDRFQ-----QPLKQLSV-QYPRQOPAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKPKPMTIITATFRSRRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLRLTYQOSS-----WHHLTNNTVTVLGLP----- 1006
Db 1222 NNHNPGRRLNQHVKVSLQKSTAVMLPKTSPALPORQSLPSHHTTKTHNPGSLPTKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGGLNLELLSKNSLIADDPREY 1046
Db 1282 FPLNPMPLPSIISKDSSTKSIISTQTAIPATTPFPASVITVETQTERSRAQTIOREQB- 1340
Qy 1047 LGQOKTAYTDGQNTTPIQ-----TPTROA--LIAFTTETTVNQSTLSAFNGSIPSDK--L 1097
Db 1341 --POKKNRTD-PNISPQDSGFTTPTAMTPPVLTAAETSV--KPSVSAFTHSPPEPNTTGI 1395
Qy 1098 STTL 1101
Db 1396 SSTI 1399

```

Search completed: February 16, 2006, 21:55:16  
Job time : 25.4867 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 21.432 Seconds  
(without alignments)  
4309.818 Million cell updates/sec

Title: US-10-754-115-47  
Perfect score: 5005  
Sequence: 1 MNIDPKLYQKTPVSVYDN.....DAEISFLTTLPLKNVKEHR 960

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2169	43.3	874	2 AI0281	probable insectici
2	2138	42.7	984	2 AE0290	insecticidal toxin
3	1833.5	36.6	952	2 AC0447	probable insectici
4	1831.5	36.6	1011	2 AD0447	probable insectici
5	308	6.2	2334	2 S32920	cell wall-associat
6	307.5	6.1	843	2 AB0539	Rhs-family protein
7	304	6.1	1404	2 E85509	hypothetical prote
8	300.5	6.0	1426	2 H64780	rhaD protein precu
9	300.5	6.0	2183	2 T37218	hypothetical prote
10	300	6.0	1404	2 E90658	RhsG core protein
11	287	5.7	1438	2 AI0093	conserved hypothet
12	285	5.7	1409	2 F91187	rhaA core protein
13	283	5.7	1397	2 C64805	rhaC protein precu
14	282.5	5.6	2167	2 AF1489	cell wall-associat
15	279.5	5.6	1512	2 AH0439	probable membrane
16	278.5	5.6	1411	2 E85145	rhaB protein precu
17	277	5.5	1377	2 C85159	rhaA protein precu
18	274	5.5	1397	2 A95570	rhaC protein in rh
19	272.5	5.4	1399	2 A99720	RhsC core protein
20	272	5.4	1317	2 F83310	conserved hypothet
21	270.5	5.4	1354	2 AG0538	Rhs-family protein
22	270	5.4	1400	2 E90886	RhsG core protein
23	267.5	5.3	1394	2 H91236	RhsH core protein
24	261.5	5.2	985	2 B86084	hypothetical prote
25	258	5.2	1398	2 B85549	hypothetical prote
26	257.5	5.1	1377	2 E86034	rhaA protein in rh
27	257	5.1	1398	2 H90698	RhsD core protein
28	247.5	4.9	1616	2 E90704	Rhs core protein w
29	242	4.8	1645	2 H85554	hypothetical prote

30	224	4.5	656	2 G85731	Rhs element associ
31	223	4.5	682	2 C64898	rhaE protein - Esc
32	221.5	4.4	456	2 AB0439	conserved hypothet
33	211	4.2	586	2 B90659	Rhs core protein [
34	211	4.2	586	2 H85509	hypothetical prote
35	211	4.2	794	2 T36972	probable membrane
36	180.5	3.6	1630	2 A53577	ascites sialoglyco
37	180	3.6	2468	2 AB3412	hypothetical prote
38	177	3.5	356	2 T37136	hypothetical prote
39	176	3.5	1959	2 AG1085	hypothetical prote
40	174.5	3.5	1417	2 H90670	probable invasiv
41	174.5	3.5	1417	2 D85521	probable adhesin e
42	174	3.5	4688	2 E92885	hypothetical prote
43	173.5	3.5	709	2 S36241	hypothetical prote
44	172.5	3.4	901	2 A49227	sialidase - Actino
45	172	3.4	4936	2 AH2515	hypothetical prote

ALIGNMENTS

RESULT 1

AI0281  
probable insecticidal toxin complex YPO2312 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AI0281  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AI0281  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-874 <KUR>  
A:Cross-references: UNIPROT:Q8ZE71; UNIPARC:UPI00000DC6A3; GB:AL590842; PIDN:CAC91117.1  
C:Genetics:  
A:Gene: YPO2312

Query Match	43.3%;	Score	2169;	DB 2;	Length	874;			
Best Local Similarity	53.0%;	Pred. No.	3.3e-115;						
Matches	451;	Conservative	106;	Mismatches	202;	Indels	92;	Gaps	15;
Qy	8	LYQKTPVSVYDNRGLIIRNIDFHRHTTANGDPDT-----RTRHQYDIHGLNQSIDPR	61						
Db	5	LFSKTPSVTVLDNRGLTVRDIAYHR-----HPDSPDVISERITHQYDARGFLTQSADPR	59						
Qy	62	LYEAKQTNNTIKPNFLWQYDLTGNPCLCTESIDAGRTVTLNDIEGRPLLTVT-----	112						
Db	60	LHGAGLM-----NFSYLTDLTGRILRTQGDNGTTVSLNDAAGRPFISASNISTSDGT	113						
Qy	113	---ATGVIQTRQYETSSLPGRLLSVAREOTPEKTSRITERLIWAGNTRAEKHNLAGOCV	169						
Db	114	EDRQAMTRTWQYEASLPGRLLSVTEQV-TGKATRIYERFYAANTRAEKSINLAGACV	172						
Qy	170	RHYDTAGVTRLESLSLTGVLSSQLLIDTQE-----ANWTGDNVTMQLNLAADIYTL	225						
Db	173	SHYDTAGLVQPDIALTCVPLSVTRLMKSADNPDAVDWQADASAWNDQDGTHTSL	232						
Qy	226	STFDATGALLTQDTAKGNIQRLAYDVAGQLNGSWTLKGQTEQVILKSLTYSAAGQKLR	285						
Db	233	TTADATGAVLTATDAKGNLQRMAYDVAGLLSGSWTLKDGTEQVIVKSLTYSAAGQKLR	292						
Qy	286	EHGNDVITEYSEVPEQRLIGIKTERP-----SDTKVLQDLRYEYDPVGNVIRINDAET	341						
Db	293	EHGNGVVTYIEYEPETQRLVGLIKTERPAGHAGAKVLQDLRYEYDPVGNVLTWDAEET	352						
Qy	342	RFWNNQKVPENTYTYDSLYQLISATGREMANIGQSHQFPSPA--LPSDNNYTYNTRYT	399						
Db	353	RFWRNQKVPENTYTYDSLYQLVSATGREMANAGQSCSLPSTTVPLPADSSAYTRYST	412						

```
Qy 400 YTYDRGNLTKI QHS SPATONNYTNI TVSNRSNRAVLSTLTEDPAQVDALFDAGHQNT 459
Db 413 YTYDEAGNLQIRNAPATNNSYTKI TVDSNRGVLSTUTENAADVDALFTAGGQTK 472
Qy 460 LISQNLNWNTRGBLQOVTLVKRDGANDREWYRISGDGRMLKINEQQAASNAQTORV 519
Db 473 LOPQHILWITARNELLKVTVPVRD-GSTDDSESYRYDAASQRIILKVSQKNTSMQTORV 531
Qy 520 TYLNLRLTONSTATTEDLQVITVGEAGRAQVRLVHWSGKPEDIDNNOLRYSYDNLI 579
Db 532 LYLPGELRSKKSGDTEGLQVITVGEAGRAQVRLVHWSGRDEITDDQIRYSYDNLA 591
Qy 580 GSSOLELDSGOIISBEEYYPYGGTALWAARNQTEASYKTIYRSGKERDAGLKYGYRY 639
Db 592 GSCLELGGDNIISABEYYPYGGTAVWARRAVEADYKTIYRSGKERDAGLKYGYRY 651
Qy 640 YQPWIGRWLSDPAGTIDGLNLYRMVRNPNVTLDPGLMPTIAERIAALKKNKVTDSAP 699
Db 652 YQPWAGRWLSADPAGWVDGLNLFWRARNPVAFIDRGNLSSELYS-QAFKRT----- 703
Qy 700 SPANATVAINIRPPVAPKPSLKASTSSQPTTHPIGAANIKP--TTSGSSIVAPLSPVG 757
Db 704 --ANKYVNIIGVRAP-----NPLGETLLKEGFPKSNFHKAKSSPTG 743
Qy 758 -----NKSTSEISLPESAQSSSSTSTNLQKSFITLYRADNRSFE----- 798
Db 744 PTAGFIADPPIYSKPSAYKQKASIDKAKALGESIDLFISKRNELIDTGNLNSLG 803
Qy 799 --EMQSKFPPEG 807
Db 804 ENRYSAPKPYG 814

RESULT 2
AE0290
insecticidal toxin [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AE0290
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AE0290
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-984 <KUR>
A;Cross-references: UNIPROT:Q8ZE10; UNIPARC:UPI00000CDBFE; GB:AL590842; PIDN:CAC91185.1;
C;Genetics:
A;Gene: sepc

Query Match 42.7%; Score 2138; DB 2; Length 984;
Best Local Similarity 47.7%; Pred. No. 2.3e-113;
Matches 473; Conservative 137; Mismatches 272; Indels 110; Gaps 24;

Qy 4 IDPKLYQKTPVSYVYDNRGLIIRNIDFHTTANGDPDT--RITRHQYDIHGHNLQSIDPR 61
Db 1 MNTSLFSKTPAVTVLDNRGLSVRSIAVHRHPSDP-DTGERITHTHQYDARGFLQTSADPR 59
Qy 62 LYEAKQTNNTIKPNFLWQYDLTGCPCTESIDAGRTVTLNDIEGRPLLTWT----- 112
Db 60 LHD-----TGRANVKYLSDLVGNALCTVSADVGTVALNDAAAGLFEVMTSVNIDTADDDL 113
Qy 113 ---ATGVIQTYQETSSLPGRLLSVAEQTPREKTSRITERLIWAGNTAEAKDHNLGACV 169
Db 114 EDRSQVNRVTWQYEGASLPGRLLSITEQVTGE--AARVTERTYAANTDEKALNLVGQCV 172
Qy 170 RHYDTAGVTRLESLSLTGTVLSQSSQLLIDTQE----ANWTGDNETVWQNMLADDIYTL 225
Db 173 SHYDTAGLRQMDSTALTGVLPSVTRRLKLDADNPDTVANMQGESSAWNDQLAADRLTTL 232
```

```
Qy 226 STFDATGALLTOTDAKGNIOKLAYDVAGLNGSMLTLKGOTEQVVIKSLTYSAAGQKLR 285
Db 233 TTADAI GAVLTITDTKGNVQRVYVNVAGLLSGSWLRVKGGAEQVIVQSLTYSAAGQKLR 292
Qy 286 EHGNVDITEYSYEPETQRLIGIKTRR----PSDTKVLQDLRYEYDPVGNVISIRNABAT 341
Db 293 EHGNVVTVTYBETQRLTGIRTSAGHASGKVLQDLRYEYDPVGNVQSIIRNABET 352
Qy 342 RFWNIQKVMENVTYDLSYQLISATGREMANIQOOSHOPSPALPSDNNYNTYTRYT 401
Db 353 RFWNIQKVPENTYVYDLSYQLVSATGRVMVNAQOGRSLPSATLPTESSAYNTYTRYT 412
Qy 402 YDRGNLTKI QHS SPATONNYTNI TVSNRSNRAVLSTLTEDPAQVDALFDAGHQNTLI 461
Db 413 YDTAGNLQIRN-TPAIGSGHTTDTITVDSNRQGVSLTITNPAEVDALFTAGGQKQLQ 471
Qy 462 SGQNLNWNTRGBLQOVTLVKRDGANDREWYRISGDGRMLKINEQQAASNAQTORV 521
Db 472 PGQHILWITARNELLKVTVPVRD-GSDDRSESYRYDGNRSQRIILKVSQKGTGSTQTORV 530
Qy 522 LPNLRLTONSTATTEDLQVITVGEAGRA--QVRLVHWSGKPEDIDNNOLRYSYDNLI 579
Db 531 LPRLELSTASGVVETTESLQIITVGEAGRAQVQVQVRLHWEKGPDAINDOLRYSYDNLI 590
Qy 580 GSSOLELDSGOIISBEEYYPYGGTALWAARNQTEASYKTIYRSGKERDAGLKYGYRY 639
Db 591 GSSTLEVDGDNVLSMEYYPYGGTAVTAHSQTEADYKTIYRSGKERDAGLKYGYRY 650
Qy 640 YQPWIGRWLSDPAGTIDGLNLYRMVRNPNVTLDPGLMPTIAERIAALKKNKVTDSAP 699
Db 651 YQPWGWRWLSADPVTVDGLNLYLMVGNPNTPSHDSNGLI-----REGQSARKLVGEAFV 705
Qy 700 SPANAT-----NVAINIRPP-----VAPKPSLKASTSSQPTTHP----- 734
Db 706 HPLHMSYFERISIBENMAMSVREAGIYITLALGEGAAAGHNIUEKTIKPSLKAVENK 765
Qy 735 IGAANIKPTTSG-----SSIVAPLSPVGNKSTSEISLPESAQ--SSSSSTSTNLQK 784
Db 766 AGALELAKSGFGTGRVGRWNASGVQGVYVNRSGEDLVYPASLQDTSNDELVNAMIKH 825
Qy 785 KSFTLYRADNRSFEEMOSKPEGKAWTP-----LDTKMARQFASIFIGQKDTSNLPKETV 840
Db 826 KIITPYTGDDYMDHII--KFNRG-KGVVPTAESAEETGVKOLINKGVAEVPARPFYEFTA 882
Qy 841 KNISTWGAQKPLKDLNLYKTKDKSVTWVSTAINTEAG-----QSSGAPLHKIDM 892
Db 883 MNVIRHG--PQVNFVFPYMWYEHDK-----VVSNDGYLVGVVARPGFPFIAMVHQGW 932
Qy 893 DLVEFAIDGOKL-----NPLPEGRTKNMV 916
Db 933 TVFD---DSKELFNFKSSNTPLFEHWQNF 961
```

```
RESULT 3
AC0447
Probable insecticidal toxin YPO3673 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C;Accession: AC0447
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A;Reference number: AB0001; MUID:21470413; PMID:11586360
A;Accession: AC0447
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-952 <KUR>
A;Cross-references: UNIPROT:Q8ZAV9; UNIPARC:UPI00000DC7E2; GB:AL590842; PIDN:CAC93143.1;
C;Genetics:
A;Gene: YPO3673
```

Query Match 36.6%; Score 1833.5; DB 2; Length 952;  
 Best Local Similarity 44.1%; Pred. No. 4.2e-96;  
 Matches 436; Conservative 138; Mismatches 294; Indels 121; Gaps 27;

QY 3 NIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRL 62  
 DB 4 SLPTQLCANTPALTTHDNRGLFAIRTLAYNRDHNETHDELISRNRYNASGQLIASRDPRL 63

QY 63 YEAKQTNNTKPNFLWQYDLTGNCPLCTESIDAGRVTTLNDIEGRPLLTATGVTQY 122  
 DB 64 E-----VDNFYQYSLSGVPLRTDSVDSGSTLQLADSAGRTVTLDAHHTRWVEY 114

QY 123 ETS--SLPGRLLSVAEQTPPEKTSRITRLIWAAGTEAKHNLKAGQCVRHVDYAGVRL 180  
 DB 115 ETGEHSL-GRPLSYHEQA-KGGLKTVTDTRFFVATNSEQDKNCNLNGQCVRHVDSAGLQAL 172

QY 181 ELSLSITGVLSQSSQLLIDTQ-EANWTGDNETVQNMMLADDIYTLSTFDATGALLTQTD 239  
 DB 173 ISQSIIGVPLQOQRELLNPKGFVDFGEKEN-WGARLSEQPFVSHSTTDALGQLLTQTD 231

QY 240 AKGNITQRLAYDVAGOLNGSWLTKGQTEQVIKSLTYSAAQOKLREHGNVDVITEYSYEP 299  
 DB 232 AKGHQRMAYNRAGOLIGSWLTKNSAEQVILRSITYSAAQOKLREESGNGVITEYRVEP 291

QY 300 ETQRLIGIKTRRPS----DTKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKVPENTY 355  
 DB 292 QTORLIGIKTRPAKDRPTR-LQDLRYDYDVPVGNILAHNDAEATRFVRNOKIVPETTY 350

QY 356 TYDSLYQLISATGREMANIQGSHQFPSPALPSDNTYNTYTRTYDGRGNLTQIHS 415  
 DB 351 RYDALYQLTEATGREATNGIQNSQLPALASLNDNQFVNYTRSYHYDRAGNLLKIQHTG 410

QY 416 PATQNNYNTNITVSNRNRVSLTSTLTPDPAQVDALPDAGGHQNTLISGONLWNTVRGELQ 475  
 DB 411 ---ASQYSTHITVSDSSNHGIIQQQGIITARDIRSQFDAGNQOQLOPQGPLRWNSRNLQ 467

QY 476 QVTLVKRDKGANDREWRYSGDGRMLKINEQQAASNAQTORVYTLNLELRLTQNSTA 535  
 DB 468 QVEPVRNDGISDS-ESYLYDGGGRVAKISLHKTHNAIQTRSVIYLAGLELRGQNDNN 526

QY 536 TTEDLOVITVGBAGRAQVRVLHWESGKPEDIDNNQIRSYDNLIGSSQLELSEQIIS 595  
 DB 527 LTESPQVITVGAAGRAQVRVLHWESGQPDVIVNDQRLYSFDFNLHLSALIELDSGDGIISQ 586

QY 596 EBYYPVGGTALWAARNQTEASVKTIRYSGKRDATGLYYGYRYQYQPIGWLSDDPAGT 655  
 DB 587 EBYYPFGGTAVLASRNTVEAKTVRYSGKRDATGLYYGYRYQYQPIGWLSADPAGT 646

QY 656 IDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINIRPPV 715  
 DB 647 IDGLNLYRMVRNPNIRWRDNNGLL--TEBQI-----NMVNLFSNI 685

QY 716 A-----PKPSLPKASTSQPHTPHIGANIIPKTTSGSIVAPLSPVGNKSTSEISLPS 769  
 DB 686 GLKNDDELKSELLKYGLSBE-PQNOIYLMNLRPMQSGS-----SSLSFPSE 732

QY 770 AQSSSSSTSTNLQKSFYLVRADNRSPFEMQSKPEGFKAWT---PLDTKMARQFASIF 826  
 DB 733 SSSSSSQTSQVD-----SGYLSVRNVH-----PFEDIKATMRPYPKQASDITTY 781

QY 827 IQQKDTSNLPKE-----TVKNISTWGAAPKPLKOLSNYIKY-TKDK-----S 866  
 DB 782 SAEDITEASPIKILIGDLTSENTPY--KSALAEG--IKYITKEVETIDPFEGGLS 837

QY 867 TVWVSTAIN-----TEAGGSSGAPLHKIDMDLYEFAIDQKLNPLPEGR 912  
 DB 838 TEQIDLTVNKLKQKOLVGIHCGAGNGRSGVIASALSIN-KQYTTD--KINSFDV--T 892

QY 913 KMWVSLLLDTQIETSSIIALNHPGVND 941  
 DB 893 HSLRGSILKDTQTYQVDTVTAKAVGIIE 921

## RESULT 4

AD0447  
 C:Species: Yersinia pestis  
 C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AD0447  
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.; Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001  
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
 A:Reference number: AB0001; MUID:21470413; PMID:11586360  
 A:Accession: AD0447  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1011 <KUR>  
 A:Cross-references: UNIPROT:Q8ZAV8; UNIPARC:UPI00000CDA64; GB:AL590842; PIDN:CAC93144.1.  
 C:GeneticS:  
 A:Gene: YPO3674

Query Match 36.6%; Score 1831.5; DB 2; Length 1011;  
 Best Local Similarity 46.0%; Pred. No. 6e-96;  
 Matches 422; Conservative 130; Mismatches 278; Indels 87; Gaps 23;

QY 1 MNIDP-KLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSID 59  
 DB 69 MPNITPDLTLCANTPTLAIHNRGFAIRTLAYNRDHNETHDELISRNRYNASGQLIASRD 128

QY 60 PRLYEAKQTNNTKPNFLWQYDLTGNCPLCTESIDAGRVTTLNDIEGRPLLTATGVTQ 119  
 DB 129 PRLE-----VDNFYQYSLSGVPLRTDSVDSGSTLQLADSAGRTVTLDAHHTRW 179

QY 120 RQYETS--SLPGRLLSVAEQTPPEKTSRITRLIWAAGTEAKHNLKAGQCVRHVDYAGV 177  
 DB 180 VEYETGEHSL-GRPLSYHEQA-KGGLKTVTDTRFFVATNSEQDKNSLNGQCVRHVDSAGL 237

QY 178 TRLSLSITGVLSQSSQLLIDTQ-EANWTGDNETVQNMMLADDIYTLSTFDATGALLT 236  
 DB 238 QALINQSIIGVPLQOQRELLNPKGFVDFGEKEN-WGARLSEQPFVSHSTTDALGQLLT 296

QY 237 QTDKAGNIQRLAYDVAGOLNGSWLTKGQTEQVIKSLTYSAAQOKLREHGNVDVITEYS 296  
 DB 297 QTDKAGHQRWAYNRAGOLIGSWLTKNSAEQVILRSITYSAAQOKLREESGNGVITEYR 356

QY 297 YEPETQRLIGIKTRRPS----DTKVLQDLRYEYDPVGNVISIRNDAEATRFWNNQKMP 352  
 DB 357 YEPQTORLIGIKTRPAKDRPTR-LQDLRYDYDVPVGNILAHNDAEATRFVRNOKIVPE 415

QY 353 NITYYDSLYQLISATGREMANIQGSHQFPSPALPSDNTYNTYTRTYDGRGNLTQI 412  
 DB 416 TTYRYDALYQLTEATGREATNGIQNSQLPALASLNDNQFVNYTRSYHYDRAGNLLK 475

QY 413 HSSPATQNNYNTNITVSNRNRVSLTSTLTPDPAQVDALPDAGGHQNTLISGONLWNTVRG 472  
 DB 476 HTG---ASQYSTHITVSDSSNHGIIQQQGIITARDIRSQFDAGNQOQLOPQGPLRWNSRN 532

QY 473 ELQOQVTLVKRDKGANDREWRYSGDGRMLKINEQQAASNAQTORVYTLNLELRLTQ 532  
 DB 533 QLQOQVEPVRNDGISDS-ESYLYDGGGRVAKISLHKTHNAIQTRSVIYLAGLELRQ 591

QY 533 STATTEDLOVITVGBAGRAQVRVLHWESGKPEDIDNNQIRSYDNLIGSSQLELSEGOI 592  
 DB 592 GNNLTDFQVITVGAAGRAQVRVLHWERGQPDVIVNDQRLYSFDFNLHLSALIELDSGD 651

QY 593 ISEEBYYPVGGTALWAARNQTEASVKTIRYSGKRDATGLYYGYRYQYQPIGWLSDDP 652  
 DB 652 ISQEEYYPFGGTAVLASRNTVEAKTVRYSGKRDATGLYYGYRYQYQPIGWLRULSADP 711

QY 653 AGTIDGLNLYRMVRNPNVTLDPDGLM--PTIAERIAALKKNKVTDSAPSANATNVAIN 710  
 DB 712 AGTIDGLNLYRMVRNPNVGLMDGDLMTDKLLAKHEANFPAKKNISSMAELKSEIEKGL- 770

QY 711 IRPPVAPKPSLPKAS-----TSSQPTHTPHIGAANIKPTTSGSS-IVAPLSPVGNKS-- 760

```

771  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db  -----LPADSKQLFLHLNGSGSDDEPSGSSG-----SSGSSEILNTSPHKIKNFH 816

761  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----TSEISLPESQSSSSSTSTNLQKSTFLYRADRNRSEEMQSPKPGFKAWTPLDYMKA 819
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db  -----FISEINLATMRP-----YKQFSSSTEDMLSE-ABRLKAYGSDITLLT 858

820  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----ROPASFIFIGKDTSNLPKETVKNISTWG-----AKPKLDLSNVIKYTKDKST 867
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db  -----LDLTSIEDPETS-I-LADKGINYIAEKQYEIDYFSEDELSENDRIVNMIKTIQNNH 917

859  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----VWVSTAINTEAGOGSSG 884
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db  -----XVGIHCAAGNGRSG 931

RESULT 5
S32920
cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text_change 09-Jul-2004
C:Accession: S32920; E69730; T47101
R:Foster, S.J.
Mol. Microbiol. 8, 299-310, 1993
A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
rotein.
A:Reference number: S32919; MUID:93302506; PMID:8316082
A:Accession: S32920
A:Molecule type: DNA
A:Residues: 1-2334 <FOS>
R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berte
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fuma, S.; Galizzi, A.; Galle
lech, J.; Harwood, C.R.; Henaut, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaite, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, J.; Sekiguchi, J.; Sekowaka, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: E69730
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2334 <KUN>
A:Cross-references: UNIPARC:UPI0000060C90; GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CA
A:Experimental source: strain 168
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai
A:Reference number: Z24350; MUID:95219088; PMID:7704263
A:Accession: T47101
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2334 <YOS>
A:Cross-references: UNIPARC:UPI0000060C90; EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PI
A:Experimental source: strain BGSCL1
C:Genetics:
A:Gene: wapa; N17G
C:Superfamily: cell wall-associated protein wapa

Query Match 6.2%; Score 308; DB 2; Length 2334;
Best Local Similarity 23.7%; Pred. No. 7.3e-09;
Matches 153; Conservative 85; Mismatches 206; Indels 202; Gaps 30;

Qy 184 SLTGTV-----LSQSSQLLIDTQBNWGTGNETWQNMADDIYTLTSTDTGALLTQ 237
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

```

```

Db 1591 SATGTWVFDDIRLIEGSLITKSTYDSN--GNVYTKBEDELG---YATSTDYDETKKTSK 1645

238  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----TDAKGNLQRLAYDAGOL-----NGSWL-----TLKGQTEQVILKSLTYS 278
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1646 TDAKGEKTTYDQADQLTNMTLSNGTSILHSYKDEGNEVSKTIRAGADQY--KFEYDV 1703

279  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  AGQ--KLREEHGNVDITEYSYEPETQRLIG-----IKTRRPSDKVLQDL 321
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1704 MGKLVKTTDPLGNVLASEYDANSNLTKTISPNNGNEVSLSYDGTDRVSKSYNGT--BKY 1760

322  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  RYEDYPGVNIVISINDAET---RFWHNQKVMPE----- 352
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1761 IFTYDKNGNETSVVNKEQNTTKRTFDNKNRLTELTDRGSGQTWTYPSDSKLKTFSWIH 1820

353  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----NTYTYVDSLYQLIS-----ATGREMANIGQSHQFPSPA 385
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1821 GDQKGTNQFTYKNLKDQMLEMKDSTSYSPFDVDENGNVQTFITG---NGGTSFSYDERN 1876

386  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  LPS-----DNNYTNNTYTRTYTDRGGLTKIQHS--SPATQNNYTTNTITVNSNRRAVLST 439
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1877 LVSSLHIGCDKNGGDLITSEYEDANGNRTTINSASGKVQVEY-----GKLNLQVKET 1929

440  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  LTEDPAQVDALFDAGGHQNTLISGQ-----NLNWNTRGBELQVVT--LVKRDGAN--- 487
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1930 -HEDGTVIEYTYDGFGRKKTVTTIKDGSSKTVNASFNIMNQLTKVNDESISYDKNGNRTS 1988

488  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  -----DDREW--YRYSGGRMLKINEQOASN---NAQTQRTY 521
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1989 DGKFTYTWDABDNLTAVTKGEDKPFATYKYDEKGNRIQKTNGKVTNYFYDGSNLNLY 2048

522  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  LPNLELRLTQNTATTEDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRAYSVD--NLIG 580
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2049 ETDADNNVTKS-----YTYGDSGL---LSYTGNGK-----KYFYHNAHG 2086

581  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  SSQLELSEGOIIEEBEYYPVGGTALMAARNQOTBSYKT---TRYSGKERD-ATGLYYY 635
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2087 DIIAISDSGTGTVAKYQV-----DAMGNPTKTEASDEVKDNRYAGYQYDEETGLYL 2140

636  | | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Qy  GYRYVQPMIGRWLSSDPAGTIDGLNL---YRMVNNPVTLLDDPG 677
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 2141 MARIYPRNGVFLSDPDGSDGSLQNGYAYGNPNPVMNVDPDG 2186

RESULT 6
AB0539
Rhs-family protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strai
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AB0539
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AB0539
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-843 <PAR>
A:Cross-references: UNIPARC:UPI000005A2A4; GB:AL513382; PIDN:CAD08754.1; PID:g16501575;
C:Genetics:
A:Gene: STV0324

Query Match 6.1%; Score 307.5; DB 2; Length 843;
Best Local Similarity 21.5%; Pred. No. 1.5e-09;
Matches 173; Conservative 110; Mismatches 257; Indels 265; Gaps 39;

Qy 5 DPKLYQKTPYTSVVDNRGLIIRNID-----FHRTTANG----- 37
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
Db 162 DP-LHQH--TVYGYDRHGQVVRITDARGGDKYLQWNEGQGLMRHTDCSGSQTANFYDERT 218
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :

```

```
QY 38 -----DPDTRITRHOYDIHGLNOSIDPRLYEAKQTNTTIKPNFLWQYDLTGMPCTES 91
Db 219 RLERTVDAENSTRYSYDNGHLTEV-----MFADGRTER-----YQFDAAGR-LVKYT 266
QY 92 IDAGR-TVTINDIEGRPLITVATGVIQTRQYETSSLPGRLLSVAEQTPB-----EKTS 144
Db 267 SPAGQITRWQRDQGRVRROTDATGRTAYEYDAY---GRLLTLTNENGESYRFYDYVLD 323
QY 145 RITER-----LIWAGNTEAEKDHNLACQCVRHRYDVTAGVTRLESLS 184
Db 324 RVTEQDGGSPRAYGYNALNAVAVIYGGEGGIRHGL-----ERDAAG-----R 370
QY 185 LTGTVLSQ-----SSQLIDTQ-----EAWTGDNETWQNMADDIYTTLSTFDATG 232
Db 371 LTAKITPTREYRYDAADRLLIEIRRRHDAEGEPEVI-----RFSYDSAG 418
QY 233 ALLTQTDAGNIQRLAYDVAGOLNGSWLTGKQTEQVLIKSLITYSAAGOKLREHGNV 292
Db 419 NLLSEETAQGVLOH-RYDVQGNRTETOMP-DGRTLRYL-----YVSGHLOQINLRDVI 471
QY 293 TEYSYE-----PETQRLIG-IKTRRPD-----TKVL-----QD- 320
Db 472 SEPTFDHLHREVQSRQGLDMRWYDRTGRLTRKLTCKMRGVVPETFDIDREYAYSGQDE 531
QY 321 -----LRYEYDPVGNVISIRNDAEATRFWNNKVPMPENTTYDSLYQLISATGR 369
Db 532 LLKKRHSRQGVTDYFYDTTGRITACRNEAYL-----DSWQYDAAANLLDRRQ 579
QY 370 EMANIGQSHQPPSPALSDNNYNTYTRT-YTYDRGGNLTKIQHSSPATQNNYNTITV 428
Db 580 ETAQAGAGS-----VWPF--NRITSYRGLHYRYDEYGRV----- 612
QY 429 SNRSNRAVLSTLTEDPAQVDALFDAGGHQNTLISQGNLNNWTRGELQOVLTKRDKGAND 488
Db 613 -----EKGRNGT-----QHYRWDAEHLTEVAVT---RGTV 642
QY 489 DREWYSGDGRMLKINEQASNNAAQRTVYLPNLRLTQNSTATTEDLQVITVGEA 548
Db 643 RRYGVYDAPGRVBEK-HELDAGKPY-NRTTFLWD-QWRLAQEC-----RL 686
QY 549 GRAQVRLHWSGKPEDI-----DNNQLRYSYDNLIGSSQLELDSQIISBEEYY 599
Db 687 GRSSSLYTYDRGSHEPLARVDRAAPGEADVLVYHTDVNGAPEMTDGGGNIWNEAGYQ 746
QY 600 PYGGTALWAARNQTEASVKTIRYSGKRD-ATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
Db 747 VWGNL---THEKETPVQONLRFQOYLDREGLHNLRYFYDPIGKFIISGDPILGAG 803
QY 659 LNLRYMVRNNPVTLDDPGLMPTIA 683
Db 804 INLYQYA-PNPLSYIDPLGLCKKFA 827

RESULT 7
85509 hypothetical protein 20268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9509)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85509
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <STO>
A:Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00000D025B; GB:AE005174; NID:g12512977; E
A:Experimental source: strain O157:H7, substrain EDL953
C:Genetics:
C:Superfamily: rhsF protein
```

```
Query Match 6.1%; Score 304; DB 2; Length 1404;
Best Local Similarity 21.0%; Pred. No. 5.4e-09;
Matches 192; Conservative 120; Mismatches 320; Indels 284; Gaps 44;

QY 18 YDNRG--LIIRNIDPHRTTANGDPDTRIT-----RHQYDIHGLNOSIDPRL--- 62
Db 646 YDAAGRITVLTNENGSSOTFRYDPVDRLTQGRFGDRTQRYHYDLTGKLTQSEDEGLVTL 705
QY 63 --YBAKQ--TWNTTK--PNFLWQYDLTGPNLCTESIDAGRTVTILN---DIEGRPLLTVA 113
Db 706 WHYDASDRITHRTVNGDPAEQWQYDEHGLWTLTSHTSEGRVSVHYGYDDKGR-----L 759
QY 114 TGVITQRYETSLPGRLLSVAEQTPBKTSRITERLIWAGNT-BAEKDHNL-----G 166
Db 760 TGERQT-----VENPE-----TGEMLWEHETGHAYSEQGLATRQEPDG 797
QY 167 QCVRHDTAGVTRLESLSLTGTVLSQSSQLLIDTQEANWTDGNETWQNMADDIYTTLS 226
Db 798 LPPVWLTYSGLYAGMKLGCTPLVEYMRDLRHRETARSFG-----GEAYELAT 846
QY 227 TFDATGALLTQTDAGNIQRLAYDVAGOLNGSWLTGKQTEQVLIKSLITYSAAGOKLREE 286
Db 847 AWNTSGQLRSR---HLNLPQDRDYDNDNGQLIRISGPQES----- 885
QY 287 HGNDVITEYSYEPETQRLIGIKTRRPSTDKVLODLRYEYDPVGNVI---SIRNDAEATRF 343
Db 886 -----REYRYS-DTGRLTGVHT---TAANLIDIDIPYATDPAGNRLPDPSTLTAT- 934
QY 344 WHNQKVPENTTYDSLYQLISATGREMANIGQSHQPPSPALSDNNYNTYNTYTYD 403
Db 935 WPDNRIRAEADAHYV--RYD-----EYGLAERTDRIPEGVI---RMHDERTHHHYD 981
QY 404 RGGNL---TKIQHSPATQNNYNTITVSNRSNRAVLSTLTEDPAQVDALFDAGGHQNTL 460
Db 982 SQHRLVFTHRLQHGEPQVESRY-----LVDPLGR--- 1011
QY 461 ISGNLNNWTRGELQOVLTKRDKGANDREWYSGDGRMLKINEQASNNAAQRTV 520
Db 1012 -TGKRV-WRRERDLTGWMSLSR---KPEETWYGDGD--RLTTVQTQTRIQTIVYQPGS 1063
QY 521 YLPNLRLTQNSTATTEDLQVITVGEAGRAQVR---VLHWESG----- 561
Db 1064 FTPLLRIE-TEN-----GEQAKARHSLAEVLQEDTGVTLPABELAVMLGRLE 1109
QY 562 ---KPEIDNNQLRYSYDNLIGSSQLELDSQIISBEEYY-----PYGG 603
Db 1110 RELRQGSVSEESQWLAQCGLTAEQMAAQLAEYIPEKHLHYCHDRGLPLALISPEGE 1169
QY 604 TALWAAR-----NOTEASY--KTIRYSGKRD-ATGLYYGYRYQYQWIGRWLSSDP 652
Db 1170 TA-WQGEYDEWGNLLGETSAQHLQOQLRPGQYDEESGLYNNRNYRYPDPLQGRYITQDP 1228
QY 653 AGTIDGLNLYMVRNNPVTLDDPGLMPTIAERTAALKKKNKVTDSAPSNATVAINR 712
Db 1229 IGLEGGWNLVQYPL-NPIEHIDPLGL-----ALDNLNYSFSDPIYKSLN----- 1273
QY 713 PPVAPKPSLPKASTSSOPTTHPIGAANIKPPTSGSSIVAPLSPVGNKSTBSISLPESAQS 772
Db 1274 -----REPTGFTVGGHG-SPTISMDDRI-----KKGSDLTIKQLASD 1310
QY 773 SSSS-----TSTNQLKKSFTLYADNRSP-----EEMQSKPPGPFKAWTPLD 815
Db 1311 IRANPKYHEGMPVVLFCSETGKGNKSAQKLANELDATVIAPDSEIIMWPDGNYA----- 1365
QY 816 TKMARQFASIFIGKD 831
Db 1366 --IMGQFARITIGKD 1379

RESULT 8
H64780
rhsD protein precursor - Escherichia coli (strain K-12)
```

C;Species: Escherichia coli  
C;Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text\_change 09-Jul-2004  
C;Accession: H64780; J50625; B30092; I69401; S16026  
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A;Title: The complete genome sequence of Escherichia coli K-12.  
A;Reference number: A64720; MUID:97426617; PMID:9278503  
A;Accession: H64780  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1426 <BLAT>  
A;Cross-references: UNIPROT:P16919; UNIPARC:UPI0000047C9A; GB:AE000156; GB:U00096; NID:9  
A;Experimental source: strain K-12, substrain MG1655  
R;Sadosky, A.B.; Gray, J.A.; Hill, C.W.  
Nucleic Acids Res. 19, 7177-7183, 1991  
A;Title: The Rhd-E subfamily of Escherichia coli K-12.  
A;Reference number: J50625; MUID:92115567; PMID:1766878  
A;Accession: J50625  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 1-378,'A',380-1166,'G',1168-1426 <SAD>  
A;Cross-references: UNIPARC:UPI00000B203B; EMBL:X60999; NID:G42732; PID:G42733  
A;Experimental source: strain K-12  
R;Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.  
J. Bacteriol. 171, 636-642, 1989  
A;Title: rhd gene family of Escherichia coli K-12.  
A;Reference number: A91901; MUID:89123133; PMID:2644231  
A;Accession: B30092  
A;Molecule type: DNA  
A;Residues: 1-100 <SAD>  
A;Cross-references: UNIPARC:UPI000016F486; GB:M21764; GB:J04224; NID:G147646; PIDN:AAA24  
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;  
J. Bacteriol. 172, 446-456, 1990  
A;Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w  
A;Reference number: I54935; MUID:90094253; PMID:2403547  
A;Accession: I69401  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1232-1426 <RSS>  
A;Cross-references: UNIPARC:UPI000000047B; GB:M29719; NID:G147644; PIDN:AAA24541.1; PID:  
C;Comment: the rhd core consist of two distinct parts: a large N-terminal core that is c  
C;Genetics:  
A;Gene: rhd  
C;Superfamily: rhaP protein  
C;Keywords: transmembrane protein  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-1426/Product: rhaD protein #status predicted <MAT>  
F;28-55/Domain: transmembrane #status predicted <TM>  
  
Query Match 6.0%; Score 300.5; DB 2; Length 1426;  
Best Local Similarity 21.7%; Pred. No. 8.8e-09;  
Matches 186; Conservative 101; Mismatches 294; Indels 275; Gaps 42;  
  
Qy 32 RTTAGDPDTRITRHOYDIGHLQSIDPRL-----YEAKQ--TWNTI--KPNFLWQYDL 82  
Db 667 RLVOQGGFGRGTQRTHYDTHGLTQSEDEGLVILWYDESRITRHTVNGEPAEQWQIDG 726  
  
Qy 83 TGNPLCTESIDAGRTVTLLN--DIEGRPLLTATGVIQTRQYETSSLPGRLLSVAEQTP 139  
Db 727 HGWLTDISHLSEGRHVAHVGYDDKGR-----LTGEQQT-----VENP 764  
  
Qy 140 EKTSTRITERLIWAGNTE-AEKDHLNAGCQVR-----HYTAGVTRLESLSLGTVLQSO 192  
Db 765 E-----TCELLWQHETKHAYNEQGLANRVTPLPPVFWLTYGSGYLAKMKLGQTP-- 816  
  
Qy 193 SSQLLIDTQEAQNTGD--NETV--WQNLADD-LYTTLSYFDATGALLTQTDAGKNIGOR 246  
Db 817 -----VEYTRDLRHRETVRSFGMAGNSAAYELTSYTPAGLQSO-----HLNS 861  
  
Qy 247 LAY--DVAGQLNGSWLTKGQTEQVIKSLTYSAAQKLRBEHGNVDITEYSYEPETQRL 304  
Db 862 LVYDRDYGWSNDGLVRISGRQ-----TREYGYG-ATGRL 896

Qy 305 IGKTRPSDKVLQDLRYEY--DPVGNVI---SIRNDAEATRFHWKQKVPENTYTYDS 359  
Db 897 ESVRTLAPD-----LDRIYATDPAGNRLPDPPELHPDSLTIT--VWPNRIADAHYVY-- 948  
  
Qy 360 LYQLISATGEMANIGQSHQFPSPALPSDNNTYNTYTRTYTYDRGNL-----TKIQHSSP 416  
Db 949 -----RHDEYGRLTETKTRIPAGVIRTDDE-----THHYHYSQHLRVFTYTRIQHGP 997  
  
Qy 417 ATQNNYNTITVNSNRV-----LSLTEDPAQVD 448  
Db 998 LVESRYLYD--PLGRMAKRVWRERDLTGWMSLSRKEPEVTWYWGWDGRLTITVQDTTRIQ 1056  
  
Qy 449 ALFDAGGHQNTLISGQNLWNTRGELQOVTLVAKDKGANDREWVRYSGDGRMLKINEQ 508  
Db 1057 TVYEPG-----SFTPLIIRVETENGERE-----KAQRSLAETLQO 1091  
  
Qy 509 QASNNAQ-----TORVTVLPNLELRLATQNSTATTEDLQVITVGBAGRAQVRVHW--ESG 561  
Db 1092 EGSNGHGVVPAELVRLDRL-----EIRADRVSSERA-----WLAQCG 1134  
  
Qy 562 KPEDIDNNQRYSDNLISSQLELDSG-----QIISSEYYPYGGTALWAAR-----NOT 613  
Db 1135 LTVQLARQVEPEYTPARKAHLVCHDRGLPLALISED-----GNTAWSAEYDEWGNOL 1188  
  
Qy 614 EAS-----YKTIRYSGKERD-ATGLYVYGVRYYOPWIGRWLSSDPAGTIDGLNLYRMVRN 667  
Db 1189 NEENPHVYQYRULPGQOHDEESGLYNNRHHYDPLQGRYITQDPMGLKGNWLYQVPL- 1247  
  
Qy 668 NPVTLLDPDGLMPTIABERIAALKKNKVTDSSAPPANATNVAINTIRPPVAPKPSLPKASTS 727  
Db 1248 NPLQOIDPMGLLQI-----WDDARSACTGGVC----- 1275  
  
Qy 728 SQPTTHPIGAIANIPTTSGSIVAPLSPVGNKSTSEISLPESAQSS-----SS 775  
Db 1276 -----GVLRSIIIGSPKFDSTADALDAL--KETQNRSLCNDMEYSGIVCKDTNGKYPAS 1327  
  
Qy 776 STTSTNLQKKSFTLYR 791  
Db 1328 KAETDNLKESYPLKR 1343  
  
RESULT 9  
T37218  
hypothetical protein SC2H4.02 - Streptomyces coelicolor  
C;Species: Streptomyces coelicolor  
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text\_change 09-Jul-2004  
C;Accession: T37218  
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, September 1998  
A;Reference number: Z21615  
A;Accession: T37218  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-2183 <OLI>  
A;Cross-references: UNIPROT:O86585; UNIPARC:UPI00000DADB3; EMBL:AL031514; PIDN:CAA20596.  
A;Experimental source: strain A3(2)  
C;Genetics:  
A;Gene: SC0EDB:SC2H4.02  
  
Query Match 6.0%; Score 300.5; DB 2; Length 2183;  
Best Local Similarity 23.0%; Pred. No. 1.8e-08;  
Matches 198; Conservative 112; Mismatches 380; Indels 171; Gaps 40;  
  
Qy 17 VYDNRGLIIRNIDFHTTANGDDPTRITRHOYDIGHLQSIDPRLYEAKQNTNITKFNF 76  
Db 1267 VTDAFGRVTERDYAGTQAG--TDYMTTRYAFDA-----ADROKSIYAHDRS 1312  
  
Qy 77 LW--QYDLTGPNPLCTESIDAGRTVTLLNIEGRPLLTATGVIQTRQYETSSLPGRLLSV 134  
Db 1313 AWTYTYDLFGQVSVTDPKGTITVTEYDALDRAVKSTDGRGEVLLFEYD---VLGRKTCM 1369  
  
Qy 135 AEQTPEKTSRITERLIWAGNTEAKDHLNLA-----GQCVRHYDPAGVTRLESLSL-TGT 188



Db 1370 WQSA---KFG-ANKLAAMFDTLAGQQDQAVRYEGGETGRAY-TQKVTRYDPLKYVNN 1424  
QY 189 --VLQSSOLL---DIOEAMWTDGN--ETVMQ-----NMLADDIYTLTTFDQATGAL 234  
Db 1425 ELTLPANDELVAAGVPARLAPSTGYNLGDTVKQAPAAVAGLSAE---TVSYTYDGLGOV 1481  
QY 235 LPOTDAKGNIOBLAYDVAGQNGSLTLKQTEQVVIKSLTYSAAGQKLREBHGNDVITE 294  
Db 1482 LTAAGTGYLQAAAYSPGLRQMTLATDPTGAKKYLNDYEAGTRRLTRSYVTDVHG 1541  
QY 295 YSEYETQRLIGIKTRPSDTKVLQDLRYEYDVGNVISIRDAEATRHQKWPENT 354  
Db 1542 F-----MLQELKYQDDAGNITSV-----SDATLLGGTGKA-DHQC 1576  
QY 355 YTYDSLYQL-----ISATGEMANIGQOSHQPSPALPSPDNNTYNTYTRYTYDR 404  
Db 1577 FYTDGHRRLSEAWTETADCSISGRTVAGLG-----GAAP-----YWTSYQYDD 1620  
QY 405 GGNLTKIQHSSPATQNNYTNITVSNRRA-VLSTLTEDPAQVDALFDAGHQNT---L 460  
Db 1621 SGLRSK-OTERHMSGDDVTVEYEGTAEGQPHALSATVTGAENASVYDETCNTETPGV 1679  
QY 461 ISGQMLNWNTRGELQVLT-LVKRDKGANDREWYRYSYSGRMLKINEQQAASNAQTQV 519  
Db 1680 RATQTLDMNAEGLAGVSEPAAGKPGATGAYVYDAGGD-----LLIRPTTDTG---ETV 1732  
QY 520 TYLPNLELRLTONSTATTEDLQVITVGEAGRAQVRLHWESKPEDIDNNQLRYSVDNLI 579  
Db 1733 LYLGTTEVHLKVSNGAAKALSGARTYKAGSAVIAVRISTAG-----VSGTKLTFLAGDHH 1788  
QY 580 GSSQLELDSGQIISSEYYPYGGTALWAARQTEASYKTIR-YSGKERD-ATGLYYGYG 637  
Db 1789 GTSGLAINADTLAFAKRWSTPFG-----APRGTAGWPDGRLGKPADATAATGLTQCA 1843  
QY 638 RYQPWIGRWLSSDPAGITD---GLNLYRMVRNVPWLLDPCLMPTTAERTAAALK--N 692  
Db 1844 ROYDEPTGRFLSVDPLEPKDPTNLNGYAYASNSPVTNSDPGSDGLGLGALGAIIG 1903  
QY 693 KYTDSAPSANATVAINIRPPVAPKPSLPKASTSSQPTTHPIGAANIKPTTSGSSIVAP 752  
Db 1904 GVGVAGVAGAAITAVG-----SLGGGGGG-----GNGTAPTSSGG----- 1941  
QY 753 LSPVGNKSTSEISLPESAQSSSTTSTNLQKSFLL-YRADNRSEEMQSKFPPEGPKAW 811  
Db 1942 -----WTQPLTKQWTFGAT--YNFITKSWDLPPNPPSPQSLEMLASMPD---W 1984  
QY 812 TPL-DTQMARQF---ASIPFG 828  
Db 1985 GIVSDPKAANRWETSRSLPFG 2005

RESULT 10  
E90658  
RhaG core protein with extension [imported] - Escherichia coli (strain O157:H7, substra  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: E90658  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A;Reference number: A99629; PMID:21156231; PMID:11258796  
A;Accession: E90658  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1404 <HAY>  
A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BA033660.1;  
C;Experimental source: strain O157:H7, substrain RMD 0509952  
C;Genetics:  
A;Gene: ECs0237  
C;Superfamily: rhesf protein

Query Match 6.0%; Score 300; DB 2; Length 1404;  
Best Local Similarity 21.4%; Pred. No. 9.1e-09;  
Matches 192; Conservative 118; Mismatches 341; Indels 246; Gaps 40;  
QY 18 YDNRG--LIIRNIDPHRTTANGDPDTRIT-----RHOYDIHGLNQSIDPRL--- 62  
Db 646 YDAAGRIITVFNENGSSQSTFRYDFVDRLTEQGFQDRTQRYHYDITGLTKTQSEDEGLVTL 705  
QY 63 --YEAKQ--TNNTIK--PNFLWQYDLTGNPLCTESIDAGRTVTLN---DIEGRPLLTWTA 113  
Db 706 WHYDASDRITHTRVNGDPAEQWQYDEHGWLTLSHTSEGRVSVHYGVDKGR-----L 759  
QY 114 TVQIVTROVYESSLPGRLLSVAEQPEEKTSRITERLWAGNT-EAEKHNLNLA-----G 166  
Db 760 TGERQT-----VENPE-----TCMLWEHETGHAYSEQGLATQEPDG 797  
QY 167 QCVRHVDTAGVTRLESLSLTGVLSSQSOLLIDTOEANNWTDGNETVWQNMMLADDIYTLIS 226  
Db 798 LPPVWLVYSGYLAKMKLGGTFLVEYMRDLRHRETARSG-----GEAYELAT 846  
QY 227 TPDATGALLTQDAKGNIQRLAYDVAGQNGSLWTLKQTEQVVIKSLTYSAAGQKLREE 286  
Db 847 AWNTSGQLRSR---HLNLPQLDRDYDNDNGQLIRISGPQES----- 885  
QY 287 HONDVITYSEYEPETORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVI---SIRDAEATRF 343  
Db 886 -----REYRYS-DTGRLTGVHT---TAANLDIDIPATDPAGNRLPDELHPDSTLTA- 934  
QY 344 WHNOKVMPENTVYDSLYQLISATGEMANIGQOSHQPSPALPSPDNNTYNTYTRYTYD 403  
Db 935 WPDNRIADAHVY--RYD-----EYGRLAEKTDRIPEGVI-----RMDERTHHYHD 981  
QY 404 RGGNL---TKIOHSSPATQNNYTNITVSNRRAV-----LSTLTEDPAQVDALF 451  
Db 982 SQHRLVFTHRIQHGSPQVESRYLYD-PLGRRRTGKVRWRERDLTGWMSLSRKEPEETWYGM 1040  
QY 452 DAGHQNTLISGQNLNWNTRGELQVTLVKRDKGANDREWYRYSYSGRMLKINEQQAAS 511  
Db 1041 D--GDRLTVOQTQTRIQTIVYQPGSFTPLLRITENGESQAKARH-----RSLAEVLQSDTG 1094  
QY 512 NNAQQRVTVLPNLELRLTONSTATTEDLQVITVG-----EAGRAQVRVLHW--- 558  
Db 1095 VTLPAELAVMLQRLERLRQGSVSESOQWLAAQGLTAEQMAQLEAGVYPERKHLHYHC 1154  
QY 559 -ESGKPEDIDNNQLRYSYDNLIGSSQ-----LELDSGQIISSEYYPYGGTALWAARQ 612  
Db 1155 DQGLPL-----GLISPGRETALAEYDEWGNLLSETSAQPL----- 1191  
QY 613 TEASYKTIYSGKERD-ATGLYYGYRYQYQWIGRWLSSDPAGTIDGLNLYRMVRNVPVT 671  
Db 1192 ----QOQLRFPQOYDEESGLYTNRNRYDPLQGRYITQDPTGLEGGWNLYQYPL-NPTE 1246  
QY 672 LLDPDGLMPTTAERIAALKKKNVTDSPSANATVAINIRPPVAPKPSLPKASTSSQPT 731  
Db 1247 HIDPGL-----ALDNLVYSPSDIYKGLNV-----REFPT 1278  
QY 732 THPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSS-----TTS 779  
Db 1279 GFTVGGHG-SPTSMSSDDRI-----KGSDLTIKQLASDIRANPKYHEGMPVVLFSCE 1329  
QY 780 TNLQKKSFTLVADNRSF-----EMQSKFPFGKAVTFLDTYMARQAPASIFIGKD 831  
Db 1330 TKGKNSPAQKLANELDATVIAPDEIWIWPDGNYA-----IMGQATARITIGKD 1379

## RESULT 11

A10093  
conserved hypothetical protein YPO0762 [imported] - Yersinia pestis (strain CO92)  
C;Species: Yersinia pestis  
C;Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: A10093  
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;



```
Db 1099 LDRLSEILADRVSBESRWLASCLGTVAQMSQMDPVYTPARKIHLHYCHDRGLPLALI 1158
QY 608 AARNOTE--ASY-----KTIYSKGRD-ATGLYYGVYVYQWIGRWL 648
Db 1159 SAEGATENCAEYDEGNLNEENPHQLQOLRLPGQQYDEESGLYNNRHYDPLHGYI 1218
QY 649 SSDPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKVKVTSAPSANATVA 708
Db 1219 TQDPITGLKGGWNYQYPL-NPVINDPQOL-----VD 1249
QY 709 INIRP-----PVAKPSLPRASTSSQPTTHPIGAANKIPTTSGSSIVAPLSPVGNKSTS 762
Db 1250 INLYPESDLIHSVADEINIPGVFT-----IGHGT- 1279
QY 763 BISLPESAQSSSSSTSTNLQKSTLYRADNRSPPEQSKPFGKAWT-PLDT-KMAR 820
Db 1280 ----PTSIESATRSIMTAK-----DLAYLIKFDGNYKDGWTLVWSCNTGKQON 1324
QY 821 QPASFIFGQKQTSNLPKSTVKNISTW---GAKPKL-----KD 854
Db 1325 SPASQAKELHTNVIGPDT---LWTWGRGTNGKLMKMDTVLTAPTNLNSKOLMAITTKD 1381
QY 855 LSNYIKY 861
Db 1382 LGNWITY 1388

RESULT 13
C64805
rhcC protein precursor [similarity] - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: C64805; 169400; T48912
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64805
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1397 <BLAT>
A:Cross-references: UNIPARC:UPI00001338AF; GB:AE000173; GB:U00096; NID:9
A:Experimental source: strain K-12, substrain MG1655
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A:Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA w
A:Reference number: 154935; MUID:90094253; PMID:2403547
A:Accession: I69400
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1221-1397 <RES>
A:Cross-references: UNIPARC:UPI000017888F; GB:M29718
A>Note: the sequence is revised in GenBank entry ECOLHSCA, release 114, (PIDN:AAC63073.1
R:Oshima, T.; Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Honjo, A.; Ikemoto, K.; Inada
T.; Mizobuchi, K.; Mori, H.; Motomura, K.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Sai
DNA Res. 3, 137-155, 1996
A:Title: A 718-Kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.7
A:Reference number: 225006; MUID:97061202; PMID:8905232
A:Accession: T48912
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1397 <OSH>
A:Cross-references: UNIPARC:UPI00001338AF; EMBL:D90709; NID:g1651305; PIDN:BA035359.1; F
A:Experimental source: strain K12; Kohara clone 174
C:Genetics:
A:Gene: rhcC
A:Map position: 15.7-16.0
C:Superfamily: rhesF protein
Query Match 5.7%; Score 283; DB 2; Length 1397;
Best Local Similarity 21.3%; Pred. No. 8.3e-08;
```

```
Matches 175; Conservative 96; Mismatches 242; Indels 308; Gaps 43;
QY 18 YDNRLIIRNIDPHRTTANGDPDTRITRHQYDIHGLNQSIDPRLYEAKQNTNTIKPNFL 77
Db 575 YDSRGQLI-----AVKDTQHETRYEYNAAGDLTVIAP-----DGSRRGT----- 615
QY 78 WQYDLTGPNLCTESIDAGRTVTLNDIEGRPLITVTATGVIQTRQVETSSLPGRLLSVAEQ 137
Db 616 -QYDAWGKRAICT-----TQGLTRSMEDYDAA---GRVIRLTSE 649
QY 138 TPEEKTSR--ITERLI-----WAGNTEAEKDHNLACQVR-----HYDTAGVTRLE 181
Db 650 NGSHTTFYDVLDRLIQETGFGRTQ-RYHDDLTKLIRSEDEGLVTHWYDEA---DRLT 706
QY 182 SLSLTGTVLSSQSLIDTQBANWGD----- 208
Db 707 HRTVNGETAERWQY-----DERGWLTDISHISEGHRVTVHYGYDSKGRLASBHLTVHHQP 761
QY 209 -NETWQNMADDIYTTLSFDTAG-ALLTQTDAGNIQRLAYDVAGOLNGSMLTKGOT 266
Db 762 TNELLMQ-----HETRHAYNAQGLANRCIPDSLPAVEWLTYG-SGWLSGMKL---GDT 810
QY 267 BOV-----IKSL-----TYSAAQOKLREEHGNVDITE----- 294
Db 811 PLVEYTRDLRHRETLRSFGRYELTAYTPAQ-LQSQHLNLSLSDRDYTWNDGELIRIS 869
QY 295 -----YSBPETQRLIGIKTRRPSDTKVLQDLRYEY--DPVGNVISIRNDAA-----A 340
Db 870 SPRQTRSYSYS--TTGRLTGVT-----TAANLDIRIPYTTDPAGNRLP--DPGLHPDSA 920
QY 341 TRFHNQKMPENTYDLSYLQISATOREMANIQQSHQPPSPALPDNNNTYTRY 400
Db 921 LSMWPDNRRIARDAHYLY-----RYDRHGR-----LTEKTDLIPGIVRTDDE--THRY 967
QY 401 TYDRGGLN---TKLOHSSPATQNNYTTNIVSNRSRAVSLTLTDPQVADLFDAGG-- 455
Db 968 HYDSOHLVHVHTRTQYAEPLVESRY-----LYDPLGR 1000
QY 456 -----HONTLISQNLNWNTRGELQVTLVKRKGANDREWYRSGDRMLKINEQ 509
Db 1001 VAKRWREERDLTG-----WMSLSRKPOVT-----WYGMWGD--RLTTI--- 1037
QY 510 ASNNAQTORVYLPNLEURLTQNSTATTEDLVITVGEAGRAQVRVLH---WESKPED- 565
Db 1038 -QNDRTIQTITYQGSFPTLRIVETAT-----GELAKTQRRSLADTLQOOSGEDGG 1087
QY 566 -----IDNNQLRYSYDNLIGSSQLSELDSEGLIISBEE-----YYP----- 600
Db 1088 SVVPPFVLVQMLDRLESILADRVSEESRRWLASCGLTVAQMSQMDPVYTPARKIHLVH 1147
QY 601 -----YGGTALWAA-----RNQTEASYKTIYSKGRD-ATGLYYGY 637
Db 1148 CDHRCGLPLALISTEGTTAWAYAEYDEWGNLLNEENPHQLQOLRLPGQQYDEESGLYNNRH 1207
QY 638 RYQOPWIGRWLSSDPAGITDGLNLYRMVRNPNVTLDDPGL 678
Db 1208 RYYPDQGRYITQDPIGLKGGWNYQYPL-NPISNIDPLGL 1247

RESULT 14
AF1489
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - I
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF1489
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, B.; Durand, L.; Dussauguet, O.; Entian, K.D.; Fsihi, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
```

A:Accession: AF1489  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-2167 <GLA>  
A:Cross-references: UNIPROT:Q92EK5; UNIPARC:UPI00000CC234; GB:AL592022; PIDN:CAC95686.1;  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: lin0454

Query Match 5.6%; Score 282.5; DB 2; Length 2167;  
Best Local Similarity 22.6%; Pred. No. 1.8e-07;  
Matches 168; Conservative 80; Mismatches 279; Indels 215; Gaps 34;

```
Qy 14 TVSVYDNRGLIIRNIDFHTTTANGDPDTRTRHGYDIHGH-----LNQSDP 60
Db 1476 TKNYDASG-----NYVTASYDEGRKTSFTFYDYGKTSDEKGNKKTLYDADN 1527
Qy 61 RLVEAKQNTNITKPNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Db 1528 ALIDTKLANGT---SVAYKID-----DNGNTTEKN-----VTASGKTQKN 1564
Qy 121 QYETSSLPGRLLSVAQTPPEKTSRITERLIWAGNTEAEKDHNLGAGCVRH-YDTAGVTR 179
Db 1565 IY-----EYVDVNKITAFTDAL-----NRTIKVEYDAAG-NE 1595
Qy 180 LESLSLTGTVLSQSOLLIDTOEAWNT-----GDNETWQNLADDI---YTTLS 226
Db 1596 TKAIMPNRGRVTESTYDSADRMGDKWINDKLAFLQYDPNGNOT----KVTDEINSIVTDK 1651
Qy 227 TFDATGALLTQTDAGKNIQRLAYDVAGOLNGSWLTGLKQOTQV-----IISLST 275
Db 1652 TYDDANRITKVAERGGDVSYTKPKYDN-----KGTDKGVEALNHGDIYAKTSYT 1705
Qy 276 YSAAGQKLRBHH-----GNDVITEYSYBETQRLIGIKTRRPSDT 315
Db 1706 YNDLDRNTRVNDGSKNAVFDFEGFNINVTAGNGTAANYTD-STQKVTNAAISSASGT 1764
Qy 316 KVLQDLAYEYDPGVNVSIRNDAEATRFWNNQKVPENTTYDLSYLQISATGREMANIG 375
Db 1765 QIL-DENYTYDAASNRTSIDNKQDG-----KTTYEYDAVNQLTKET----- 1804
Qy 376 QSQHQPSPALPSDNNTYTNTRYTYDRGNGNLTKIQHSSPATQNNYTNITVNSNRNA 435
Db 1805 -----LPDGT-----VKAYTYDGFGRNTQV-----AISGSEKTIIDASYNDGNQ 1843
Qy 436 VLSTLTEDPAQVDAL-FDAGGHQNTLISGQNLNNTNTRGELQOQVTLVKRDKGANDREWYR 494
Db 1844 LVSWNGE-----ALTYDANGNR-TSDGKYTYTWDTGRLSSIT-----KKGSEBPFTSYT 1892
Qy 495 YSGDGRMLKINEQOASN---NAQOTQVYLPNLELRLTQNSTATTEDLQVITVGEAGRA 551
Db 1893 YDDNRRLSKTVGVTNNYHYDGSIDVLY-----ETDGGKVVRQVYVYSD 1939
Qy 552 QVRVLWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISBEEYYPYGGTALWAARN 611
Db 1940 NVRLAKQNGK-----TLYHY-NAGDVIALTDEAGKLVAYEYDANGVNLKVTAST 1991
Qy 612 QTEASYKTIRYSGKRD-ATGLYYGYRYYPQFWIGRWLSSDP-AGTIDG-----LNLRYMVR 666
Db 1992 E-EAKANPYGVAGTYDKIEIQYLMARYEPYEQGVFTAYDFYDGEDDDPQTMMNGYNYAN 2050
Qy 667 NNPVTLDDPGLMPTIARIIAA 688
Db 2051 NNPVMMFDPDG---NVAWIIAA 2069
```

## RESULT 15

AH0439  
probable membrane protein YPO3615 [imported] - Yersinia pestis (strain CO92)  
C:Species: Yersinia pestis  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
C:Accession: AH0439  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AH0439  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1512 <KUR>  
A:Cross-references: UNIPROT:Q8ZB14; UNIPARC:UPI000000CDA53; GB:AL590842; PIDN:CAC93084.1;  
C:Genetics:  
A:Gene: YPO3615

Query Match 5.6%; Score 279.5; DB 2; Length 1512;  
Best Local Similarity 22.8%; Pred. No. 1.5e-07;  
Matches 179; Conservative 106; Mismatches 300; Indels 201; Gaps 43;

```
Qy 9 YQKTPYTSV-YDNRGLIIRNIDFHTTTANGDPDTRTRHGYDIHGHNLQSDP----- 60
Db 677 YPSQOTWSFRYKKGVELRKI-----IAEQVWRVYHDHGHCLSTIIDPKGNSTAV 727
Qy 61 -----RLYE-----AKQNTNITKPNFLWQ---YDLTGPNLCTES 91
Db 728 TLDVGLRGLFSHQNALGELTRYTHSDAHASPAQSVTKMMPDGVQAIAYD-SEKRIAAIT 786
Qy 92 IDAGRTVTLNDIEGRPLLT--VTATGVIQTRQYETSSLPGRLLSVAEQTPPEKTSRITER 149
Db 787 DGAKT-TRYEYGGFDLLTGLIRPDGQRLTFGYDITRLNQVWTWASGDTYRTYTRDR---- 841
Qy 150 LIWAGNTEAEKDHNLGAGCVRH-YDTAG---VTRLESLSLTGTVLSQSSQLLIDTOEANW 205
Db 842 ---AGQVISETD--FTGRTVHYQYDAVGRIRGARYPQORLVRWHVYSMDQDVL---AQQTW 893
Qy 206 TGDNETWQNLADDIYTTLS-TFDATGALLTQTDAGKNTQRLAYDVAGQ-----LNG-- 257
Db 894 HCD-----ALSSLTGTVSYGYDGAGRLLSATNADAVVE-FDYDEAGQVLAERLNGRE 945
Qy 258 ---SMLTLKQ--TEQVIKSLT--YSAAGQKLRBHHGNDVITEYSYBETQRLIGIKTR 310
Db 946 VRHQWDALNGTPVARQVEGLTFTVGAQGLTLQLAGHQPLQLQHD-----RLGRET 1000
Qy 311 RPSDTKVLQDLRYEYDPVG---NVISIRNDAEATRFWNNQKVPEN-----TITY 357
Db 1001 RESAAGFIQ--ACNYTFSGLLAHQAAGNSA---LFQOQLIAEPSPALHGSVAVNRSQY 1054
Qy 358 DSLYQLISATGREMANIQSQSHQP--SPALPSDNNTYTNTRYTYTYDRGNGNLTKIQHSS 415
Db 1055 DRATNVV---GMDDGRWGTQYQYDRNDQVVRADFGGFLPLQEQFSYDVNQNRE-HRCL 1110
Qy 416 PATQNNYTNITVNSNRNAV-----LSTLTEDPAQVDALFDAGGHQNTLISGQ 464
Db 1111 PRGAQVLAQASQOQOQAGRVVRKGDQSVRYDAAAGRLVEKRSQKQ-----GYRQL---W 1161
Qy 465 NLNNTNTRGELQOQVTLVKRDKGANDREW-YRYSGDGRRMLKINEQOASNNNAQOTQVYLP 523
Db 1162 RYRNQEQDQLSELL--TPTGA---RWRYGYDAFGRIRKL-----RVVDTP 1202
Qy 524 NL-ELRLTQNSTATT-----EDLOVITVGEAGRAQVRLHW----- 558
Db 1203 PLNEMDAPSTGPATASLAGYAVLWSGDLIEVEVPYADGTV--AYEQGIHHLWYAFGGTLTP 1260
Qy 559 ----ESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISBEEYYPYGGTALW-AARNQT 613
Db 1261 MARYAQKQ-----LHYVVADHLGTPRELNLNQGKVVWASRLSTWGQAEWLWQAANEE 1312
Qy 614 EASYKTIYRSGKERDA-TGLYYGYRYYPQFWIGRWLSSDPAGTIDGLNLYRMVRNPNVTL 672
Db 1313 DRVSCNLRPAQYADAEGLHYNFRYYDGTGQVLCPPDPIGLEGLNPNYGVV-HNPVSW 1371
Qy 673 LDDPGL 678
Db 1372 VDPLGL 1377
```

Search completed: February 16, 2006, 21:45:04  
Job time : 26.432 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 128.785 Seconds  
(without alignments)  
5259.201 Million cell updates/sec

Title: US-10-754-115-47

Perfect score: 5005

Sequence: 1 MKNIDPKLYQKTPVSVYDN.....DAEISFLTLTKNVKPKHR 960

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5005	100.0	960	2	Q8GF97 PHOLU
2	3300.5	65.9	970	2	Q7N7Y4 PHOLU
3	2879	57.5	938	2	Q8GF91 PHOLU
4	2850	57.0	949	2	Q8GF45 PHOLU
5	2833.5	56.6	938	2	Q7N7Y7 PHOLL
6	2821.5	56.4	959	2	Q7N7X5 PHOLL
7	2768.5	55.3	965	2	Q7MZU2 PHOLL
8	2740	54.8	1043	2	Q7MZV7 PHOLL
9	2740	54.7	1043	2	Q85157 PHOLL
10	2539	50.7	915	2	Q7N7Z1 PHOLL
11	2535.5	50.7	915	2	Q33EP1 PHOLL
12	2386	47.7	936	2	Q7MZ20 PHOLL
13	2216	44.3	1016	2	Q6Q1Y7 XENNE
14	2184	43.6	973	2	Q9F920 XENNR
15	2173	43.4	994	2	Q66A99 YERPS
16	2169	43.3	874	2	Q8ZE71 YERPE
17	2168.5	43.3	978	2	Q6AP55 YERFR
18	2148	42.9	982	2	Q66A38 YERPS
19	2138	42.7	984	2	Q8ZE10 YERPE
20	1990.5	39.8	1014	2	Q33RN9 XENNE
21	1833.5	36.6	952	2	Q8ZAV9 YERPE
22	1831.5	36.6	943	2	Q8D1P5 YERPE
23	1831.5	36.6	1011	2	Q8ZAV8 YERPE
24	1816	36.3	952	2	Q665G1 YERPS
25	1763	35.2	1045	2	Q693A0 YEREN
26	1301	26.0	886	2	Q87X45 PSESM
27	1274	25.5	927	2	Q4ZX15 PSESY
28	1268.5	25.3	940	2	Q87X48 PSESM
29	1267	25.3	956	2	Q4ZYW6 PSESY
30	1266.5	25.3	920	2	Q4ZP60 PSESY
31	1238.5	24.7	923	2	Q4ZTS1 PSESY

32 1235.5 24.7 881 2 Q4ZP54 PSESY Q4ZP54 pseudomonas  
33 1223.5 24.4 955 2 Q4ZP55 PSESY Q4ZP55 pseudomonas  
34 1217.5 24.3 939 2 Q4ZP59 PSESY Q4ZP59 pseudomonas  
35 1215.5 24.3 922 2 Q887R0 PSESM Q887R0 pseudomonas  
36 618 12.3 163 2 Q8WCC0 XENNR Q8WCC0 serratia pr  
37 590.5 11.8 2439 2 Q4HWU2 GIBZE Q4HWU2 gibberella  
38 518 10.3 1140 2 Q73L45 TREDE Q73L45 treponema d  
39 517.5 10.3 1488 2 Q73L43 TREDE Q73L43 treponema d  
40 492 9.8 1126 2 Q73MB8 TREDE Q73MB8 treponema d  
41 479.5 9.6 3320 2 Q73MFO TREDE Q73MFO treponema d  
42 470 9.4 103 2 Q8KSP1 XENNE Q8KSP1 xenorhabdus  
43 395.5 7.9 1746 2 Q5GZ16 XANOR Q5GZ16 xanthomonas  
44 357.5 7.1 1513 2 Q7NY44 CHRVO Q7NY44 chromobacce  
45 354.5 7.1 1551 2 Q4LW78 BURK Q4LW78 burkholderi

#### ALIGNMENTS

##### RESULT 1

Q8GF97 PHOLU PRELIMINARY; PRT; 960 AA.  
AC Q8GF97;  
DT 01-MAR-2003 (Tremblrel. 23, Created)  
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)  
DE TccC3.  
GN Name=TccC3;  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_TaxID=29488;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=22454990; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;  
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,  
RA fFrench-Constant R.H.;  
RA "The tc genes of Photorhabdus: a growing family.";  
RL Trends Microbiol. 9:185-191(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=W14;  
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;  
RA Waterfield N.R., Daborn P.J., fFrench-Constant R.H.;  
RA "Genomic islands in Photorhabdus.";  
RL Trends Microbiol. 10:541-545(2002).  
DR EMBL; AF346500; AAO17204.1; -; Genomic\_DNA.  
DR InterPro; IPR002345; Lipocalin.  
DR InterPro; IPR006530; YD.  
DR Pfam; PF05593; RHS repeat; 4.  
DR TIGRFAMs; TIGR01643; YD repeat 2x; 1.  
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.  
SQ SEQUENCE 960 AA; 107061 MW; 595DE383870C8E9E CRC64;

Query Match 100.0%; Score 5005; DB 2; Length 960;  
Best Local Similarity 100.0%; Pred. No. 3.1e-259;  
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHQYDIHGLNQSIDP 60  
|||||  
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHQYDIHGLNQSIDP 60  
|||||  
Qy 61 RLYEAKQTNNTIKNFELWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIOFR 120  
|||||  
Db 61 RLYEAKQTNNTIKNFELWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIOFR 120  
|||||  
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLACQCVRHVYDTAGVTRL 180  
|||||  
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLACQCVRHVYDTAGVTRL 180  
|||||  
Qy 181 ESSLTGTVLVSQSSQLLIDTQEAQNWGTQVWQNLADDIYTTLTSTFDATGALLTQIDA 240  
|||||

```
Db 181 ESLSLGTGVLSSQSLLDITQEANWTDGNETVWQNLADDITVTLSTFDATGALLTQTD 240
Qy 241 KGNLQRLAYDVAGQINGSWLTKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Db 241 KGNLQRLAYDVAGQINGSWLTKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Qy 301 TORLIGIKTRPSPDKVLQDLRYEYDVPVGNVISIRNDAEATRFVHWNKQWPNENTYTDLS 360
Db 301 TORLIGIKTRPSPDKVLQDLRYEYDVPVGNVISIRNDAEATRFVHWNKQWPNENTYTDLS 360
Qy 361 YQLISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSSPATON 420
Db 361 YQLISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSSPATON 420
Qy 421 NYTTNITVSNRNSRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGLQOVTLY 480
Db 421 NYTTNITVSNRNSRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGLQOVTLY 480
Qy 481 KRDSGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTATTEDL 540
Db 481 KRDSGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTATTEDL 540
Qy 541 QVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESEBYYP 600
Db 541 QVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESEBYYP 600
Qy 601 YGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGTIDGLN 660
Db 601 YGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGTIDGLN 660
Qy 661 LYRVRNPNVTLDPDGLMPTIAPRIAAALKKQKVTDSAPSNATVAINRPPVAPKPS 720
Db 661 LYRVRNPNVTLDPDGLMPTIAPRIAAALKKQKVTDSAPSNATVAINRPPVAPKPS 720
Qy 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTST 780
Db 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTST 780
Qy 781 NLQKSFLLYPADNRSFEEMQSKPEPGKAWPLDITKWARQFASIFIGOKDTSNLPKETV 840
Db 781 NLQKSFLLYPADNRSFEEMQSKPEPGKAWPLDITKWARQFASIFIGOKDTSNLPKETV 840
Qy 841 KNIISTWGAQPKLKDLSNYIKTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
Db 841 KNIISTWGAQPKLKDLSNYIKTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
Qy 901 GQKLNPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLLTIPKKNVQPKHR 960
Db 901 GQKLNPLPEGRTKNMPVSLLLDTPQIETSSIIALNHGVPNDABISFLLTIPKKNVQPKHR 960
```

## RESULT 2

```
Q7N7Y4 PHOLL
ID Q7N7Y4 PHOLL PRELIMINARY; PRT; 970 AA.
AC Q7N7Y4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein Tccc3.
GN Name=tccc3; OrderedLocusNames=plu0967;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
EX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt.886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derosé R., Derzelle S., Freyssié G., Gaudriault S.,
RA Médigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
```

```
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CAE13262.1; -, Genomic_DNA.
DR Photolust; plu0967; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS repeat; 5.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 1.
KW Complete proteome.
SQ SEQUENCE 970 AA; 109550 MW; C62D47037EAD644A CRC64;

Query Match 65.9%; Score 3300.5; DB 2; Length 970;
Best Local Similarity 67.6%; Pred. No. 5.5e-168;
Matches 679; Conservative 81; Mismatches 151; Indels 93; Gaps 17;

Qy 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPDTRITRPHQYDIGHLNQSIDP 60
Db 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPDTRITRPHQYDIGHLNQSIDP 60
Qy 61 RLYEAKOTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Db 61 RLYEAKOTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVRRHYDTAGVTRL 180
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVRRHYDTAGVTRL 180
Qy 181 ESSLTGTVLSSQSLLDITQEANWTDGNETVWQNLADDITVTLSTFDATGALLTQTD 240
Db 181 ESSLTGTVLSSQSLLDITQEANWTDGNETVWQNLADDITVTLSTFDATGALLTQTD 240
Qy 241 KGNLQRLAYDVAGQINGSWLTKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Db 241 KGNLQRLAYDVAGQINGSWLTKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Qy 301 TORLIGIKTRR-----PSDTKVQLDLYEYDVPVGNVISIRNDAEATRFVHWNKQWPNENTY 355
Db 301 TORLIGIKTRRSLDTPNPSDTKVQLDLYEYDVPVGNVISIRNDAEATRFVHWNKQWPNENTY 355
Qy 356 TYDSLQYLIISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSS 415
Db 356 TYDSLQYLIISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSS 415
Qy 361 IYDSLQYLIISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSS 420
Db 361 IYDSLQYLIISATGREMANIQOSSHQSPSPALPSDNNNTYTRTYTYDRGGLTKIQHSS 420
Qy 416 PATQNNYTNITVSNRNSRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGL 475
Db 416 PATQNNYTNITVSNRNSRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGL 475
Qy 421 PATQNNYTNITVSHQSNRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGL 480
Db 421 PATQNNYTNITVSHQSNRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWNTRFGL 480
Qy 476 QVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 535
Db 476 QVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 535
Qy 481 QVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 540
Db 481 QVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 540
Qy 536 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESE 595
Db 536 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESE 595
Qy 541 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESE 600
Db 541 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSDNLIIGSSQLELDSGQIIESE 600
Qy 596 EBYYPYGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGT 655
Db 596 EBYYPYGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGT 655
Qy 601 EBYYPYGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGT 660
Db 601 EBYYPYGGTALWAARNQTEASKTIRYSKERDATGLYYGYRYQYPWIGRWLSSDPAGT 660
Qy 656 IDGLNLYRVRNPNVTLDPDGLMPTIAPRIAAALKKQKVTDS-----RIAAKKNKVTDS 697
Db 656 IDGLNLYRVRNPNVTLDPDGLMPTIAPRIAAALKKQKVTDS-----RIAAKKNKVTDS 697
Qy 661 IDGLNLYRVRNPNVTKYHDTDGRAPEIKDVEHTVIYGFSHRSGSEMKGAFANKNPVT-- 718
Db 661 IDGLNLYRVRNPNVTKYHDTDGRAPEIKDVEHTVIYGFSHRSGSEMKGAFANKNPVT-- 718
Qy 698 APSANATVAINRPPVAPKPSLPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPV 757
Db 698 APSANATVAINRPPVAPKPSLPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPV 757
Qy 719 ----IDENAGLGMGSLSDYFRISQNLKEDSPPKLNEENIR-TNSSKYLVA-----MG 769
Db 719 ----IDENAGLGMGSLSDYFRISQNLKEDSPPKLNEENIR-TNSSKYLVA-----MG 769
Qy 758 -NKSTSEISLPESAQSSSSSTTSLNOKKSFLLYPADNRSFEEMQSKPEPGKAWPLDIT 805
Db 758 -NKSTSEISLPESAQSSSSSTTSLNOKKSFLLYPADNRSFEEMQSKPEPGKAWPLDIT 805
Qy 770 IEESATDKIKQEAQEYQYAVTA-----SWSFHWENKGLDQSKVKNLLPSKSLPSG 823
Db 770 IEESATDKIKQEAQEYQYAVTA-----SWSFHWENKGLDQSKVKNLLPSKSLPSG 823
Qy 806 -EGKAWTPLDITKWARQFASIFIGOKDTSNLPKETVKNKIST-WGAKPKLKDLSNYIKTK 863
Db 806 -EGKAWTPLDITKWARQFASIFIGOKDTSNLPKETVKNKIST-WGAKPKLKDLSNYIKTK 863
```



```

Db 824 LSGQRFOKLD-----AFILKNSDGSDFTIQORQDILTKFASAGNKGLESTLS-AA 874
Qy 864 DKSTVWSTAINTEAGSGSSAPLHKIDMDLY-----EF-----AIDGQKL----- 904
Db 875 DQS--WLKNTATAPRQTS-----KLGWDWFIKQLASPDFRFVWAGYNGEKLITTEQLKS 927
Qy 905 -NLPPEGKTKMVPSELDTPOIETSSIIALNHGVPNDABEISFL 947
Db 928 EOPWMTGKRREKGP--EYAEAITFSEIRHAYRKKYDSTINFI 968

RESULT 3
Q8GF91 PHOLU PRELIMINARY; PRT; 938 AA.
AC Q8GF91;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE TccC5.
GN Name=TccC5;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_taxid=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AA017210.1; -; Genomic_DNA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs_repeat; 4.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
DR SEQUENCE 938 AA; 105094 MW; 080127894806940B CRC64;

Query Match 57.5%; Score 2879; DB 2; Length 938;
Best Local Similarity 60.9%; Pred. No. 1.9e-145;
Matches 591; Conservative 90; Mismatches 186; Indels 104; Gaps 16;

Qy 1 MKNIDPKLYQKTPVSVVDNRLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Db 1 MENIDPKLYHTPTVSVVDNRLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Qy 61 RLYEAKQNTNTIKPNFLWOYDITGNPLCTESIDAGRTVLNDIEGRPLITVATGVIQTR 120
Db 61 RLYDAKQTNNAVQFNIWHNLTGNILATESVDAGRTITLNDIEGRPLITVATGVIQTR 120
Qy 121 QYETSLLPGRLLSVAEQ-TPEKTSRITERLWAGNTEAEKHNLAGQCVRHYDTAGVTR 179
Db 121 RYEDNTLPGRLIAISEQQAEBK--TERLIWAGNTPQKHNLAGQCVRHYDTAGLTQ 177
Qy 180 LESLSLTGTVLSQSQQLLIDTQEAQNTGNETVWQMLADDIYTLSTFDTGALLTQTD 239
Db 178 LNSLALTGAVLSQSQQLLTDNODADTGBDQSLQKLSDDVYITQSNLTGALLTQTD 237
Qy 240 AKGNLQRLAYDVAGOLGNSWLTGKQTEQVIKSLTYSAAQOKLREHGNNDVITEYSYEP 299
Db 238 AKGNLQRLAYDVAGOLGNSWLTGKQTEQVIKSLTYSAAQOKLREHGNNDVITEYSYEP 297
Qy 300 ETQRLIGIKTRRPSDKVLQDLRYQDYPGVNVISIRNDAEATFRFHNOKVAPENSTYDTS 359

```

```

Db 298 ETQRLIGIKTRRPSDKVLQDLRYQDYPGVNVISIRNDAEATFRFHNOKVAPENSTYDTS 357
Qy 360 LYQLISATGRMANIGQSHQPPSPALPSDNNNTYNTYTRTYDTRGGNLTKEHSSPAFQ 419
Db 358 LYQLISATGRMANIGQSHQPPSPALPSDNNNTYNTYTRTYDTRGGNLTKEHSSPAFQ 417
Qy 420 NNYTNIITVSNRNRVAVLTETEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOVTL 479
Db 418 NNYTTDITVSNRNRVAVLTETEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOVTL 474
Qy 480 VKRDKGANDD--REWYRSGDGRRLMKINEQOASNNNAQTORVYLPNLELRLTQNSTATT 537
Db 475 -----ANNSAGNEWIRYDSNGIRQLKVNEQOTQNIPOQORVYLPGLIEIRTNQNTATT 528
Qy 538 EDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLGSSQLDSDGQIISEE 597
Db 529 EELHVITLKGAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLGSSQLDSDGQIISEE 588
Qy 598 YYPGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQPWIGRWLSSDAGTTID 657
Db 589 YYPGGTALWAARNQTEASYKTIYRSGKRDVTLGLYYGYRYQPWIGRWLSSDAGTTID 648
Qy 658 GLNLYRMVRNPNVTLLDPDGLMPTIAERIAALKKKNKVTDSAPSANATNVAI-----N 710
Db 649 GLNLYRMVRNPNVTQFDVQGL-----SPANRTEEALIKQGSFTG 687
Qy 711 IRPPVAPKPSLPK-----ASTSQP-----TTHPIGAANIKP-----TTSRSSIVAPL 753
Db 688 MEEAVYKQWAKPQTFKQRAIAAQAQTEQAESLTNPP--SVDISPIKNYTTDSSQINAAI 745
Qy 754 SPVGNKSTSEISLPESAQSSSSTSTNLQKKSFTLYRADNRSEFEMQSKPFEGKAWTP 813
Db 746 RE--NRITPAV---ESLDATLSSLDQRQWRTYRVMTYVDNS-----TP 784
Qy 814 LDTKWARQFASIFIGQKDTSNLPKETVKNISWGAQPKLKDLSNVIKYTKDKSTVWVSTA 873
Db 785 SPWHSQPGNSINVGDIIVSDNAYLSTSAH-----RGFLNFVHKHKTSETRYVKNWA 834
Qy 874 INTAGGQSSGAPLHKIDMDLYEFAID-----GQKLNPLPEGRTKMVPSELDTDPQ 925
Db 835 FLTNAGVNVNPAASMYNNAGEQVFKMDLNDRSKSLAEKLRVSGPQSQABILLPRETQ 894
Qy 926 IETSIIALNH 936
Db 895 FE---VWSMKH 902

RESULT 4
Q8GFAS PHOLU PRELIMINARY; PRT; 949 AA.
AC Q8GFAS;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE TccC4.
GN Name=TccC4;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_taxid=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;

```



```
Db 418 NNYTTNITVSNYRNLAVLSTLTEDPTQVDALFSGGHQTNLSGQVLWTTPRGELKQV-- 475
Qy 480 VKRDGANDREWYSGDGRMLKINEQOASNNAAQTQRTVYLPNLELRLTQNSTATTED 539
Db 476 -----NSSAGNWHYDSNGTQLKVNQEQONIAQQQRTVYLPGLRLTTHQGSTTTEY 530
Qy 540 LQVITVGBAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEERY 599
Db 531 LQVITLKGAGRAQVRLHWESGKPEDINNQLRYSYDNLIGSSQLELSEGOIIEEERY 590
Qy 600 PYGGTALWAARNQTEASVKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 659
Db 591 PFGGTALWAARNQTEASVKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 650
Qy 660 NLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAI-----NIR 712
Db 651 NLYRMVRNPNVTFQVQGL-----SPANTEEALIKQSFTGME 689
Qy 713 PPVAPKPSLPKA-----STSSQPTTHPIGAAN-----IKPTTSGSSIVAPLSPVGN 758
Db 690 EAVYKMAKQPTFKRQRAIATQEQEAHQLLTNNPGVDTSPDKYTTDSSQI-----N 742
Qy 759 KSTSE---ISLPESAQSSSTSTNLQKSTFLRADNRSEENQSKFPEGKAWTFLD 815
Db 743 TAIRENRITSIVKOLDSSLSALQDRQIRVTVRVMTYIDN-----SK-----ESP 786
Qy 816 TKMARQFASIFIGQKDTGNLKPETVKNISTWGAQKPLDLSNYIKYTKDKSTVMVSTAIN 875
Db 787 WHSPQEGNSINVDIVSNAYLSTSAH-----RGFLNFVHKKTSETRYVKMAFL 836
Qy 876 TEAG-----GSSGAPLHKIDM-DLYEPAIDGQKL--NLPPEGRTKMWPSLLD 923
Db 837 TWTGVNAAKSKYNNENAEKIFKMLDSDWSKSVFEKLKIRANGPAGQAEILFPR---ET 893
Qy 924 POIETSSIIALHGPVNDASFLTIPLKVKVP 957
Db 894 P-----FVVSVMRHQRD-----TYVLLQDIKP 916

RESULT 6
Q7N7X5 PHOLL PRELIMINARY; PRT; 959 AA.
AC Q7N7X5;
ID Q7N7X5;
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC4.
GN Name=tccC4; OrderedLocuNames=plu0976;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Sigurinet P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; EX571862; CAB13271.1; -; Genomic_DNA.
DR PhotoList; plu0976; -.
DR InterPro; IPR002345; Lipocalin.
DR DR Pfam; PF05593; Rhs repeat; 2.
DR TIGRFam; TIGR01643; YD repeat_2x; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Complete proteome.
```

```
SQ SEQUENCE 959 AA; 108232 MW; 1482133DBB080EB7 CRC64;
Query Match 56.4%; Score 2821.5; DB 2; Length 959;
Best Local Similarity 67.0%; Pred. No. 2.4e-142;
Matches 550; Conservative 74; Mismatches 132; Indels 65; Gaps 7;
Qy 1 MKNIDPKLYOKTPTVSYVDNREGLIIRNIDFRTTANGDPDTRITRHOYDIHGLHNSIDP 60
Db 1 MKNIDPKLYOHTPTISVYVDNREGLIRNIDFRTSAGGDTDRITRHOYDVHGLHSQSIDP 60
Qy 61 RLYBAKOTNNIKPNFNLWOYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Db 61 RLYDAKQLDNSINPNFLWQNLTGTLTSEADAGRTVALNDIEGRVLTATNAIQT 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGCVRYHYDTAGVTRL 180
Db 121 QYEANTLPGRLLSVSEQAAGEKIPLVTEHFIWAGNTQTEKDHNLGAGCVRYHYDTAGVTRL 180
Qy 181 ESLSLTGTVLSSQQLLIDTQEAHWGNTGNETVWQNMADDIYTTLTSTEDATGALLTOTDA 240
Db 181 KSLSLTENILSQSRQLLADGQQAOWTGNNDPLMQKLSSEIYTTQSTFDATGALLTOTDA 240
Qy 241 KGNIQRLAYDVAGOLNGSWLTLKGQTEQVVIKSLTYSAGQKLRREHNDVITEYSYPE 300
Db 241 KSNIQRLAYVAGQLKGSWLTILKQSEQVVIKSLTYSAGQKLRREHNGALTEYSYPE 300
Qy 301 TORLIGIKTRPRSPDKVLQDLRYEYDPVGNVISIRNDAEATRFMHNKQMPENTVYDGL 360
Db 301 TLRLLIGTTTRQSDKNVLQDLRYEYHDPVGNIIIRNDAEATRFMHNKQMPENVYDGL 360
Qy 361 YQLISATGREMANIQOSSHQFPSP-ALPSDNNNTYNTYTRTYDRGGNLTKIOHSSPATQ 419
Db 361 YQLISATGREMANIQOQNRPSPPLPTDNNNTYNTYTRTYDRGGNLAQIRHSSPASQ 420
Qy 420 NNYTTNITVSNRNLAVLSTLTEDPAQVDALFSGGHQTNLSGQVLWTTPRGELKQVTL 479
Db 421 NNYTTDITVSNRNLAVLSTLTEDPTQVDALFSGGHQTNLSGQVLWTTPRGELKQVTP 480
Qy 480 VKRDGANDREWYSGDGRMLKINEQOASNNAAQTQRTVYLPNLELRLTQNSTATTED 539
Db 481 VNREN--TSDREWYRYGNDGMRLKVNQEQNSTQQRQVTVLPGLRLTTHQGSTTSED 538
Qy 540 LQVITVGBAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEERY 599
Db 539 LHVITVGAAGRAQVRLHWATTPPAGISNNQLRYSYDNLIGSSQLELNEGOIISQEEY 598
Qy 600 PYGGTALWAARNQTEASVKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 659
Db 599 PFGGTALWAARNQTEASVKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 658
Qy 660 NLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAIIRPPVAPKP 719
Db 659 NLYRMVRNPNITLYDNDGLAP-----PRVRPRH 686
Qy 720 SLPKASTSSQ-----PTTH---PIGANIKPTT-----SGSSIVAPLSPVG 757
Db 687 PDSEGESEDESSIGYRALRSDHEPSIHGLPPPEGANLNISAYAHVRACTSAKVSKSWIS 746
Qy 758 NKSTSEISLPESAQSSSTSTNLQKSTFLYRADNRSE 798
Db 747 YSRSLKVAASWASGSGSRVVKFRKK-----NNRSFD 779

RESULT 7
Q7MZU2 PHOLL PRELIMINARY; PRT; 965 AA.
AC Q7MZU2;
ID Q7MZU2;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC6.
GN Name=tccC6; OrderedLocuNames=plu4182;
OS Photorhabdus luminescens (subsp. laumondii).
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbr886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Tachout S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; BX571873; CAE16554.1; -; Genomic_DNA.
DR Photolista; plu4182; -.
DR InterPro; IPR002345; Lipocalin.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 965 AA; 109741 MW; A27BC9F13C5822F1 CRC64;

Query Match 55.3%; Score 2768.5; DB 2; Length 965;
Best Local Similarity 64.0%; Pred. No. 1.6e-139;
Matches 541; Conservative 96; Mismatches 165; Indels 43; Gaps 5;

Qy 1 MKNIDPKLYQKTPVSVVYDNRGLIIRNIDFHRITANGDPDTRITRHHQYDIGHLNQSIDP 60
Db 1 MNNFDPKIYQHTPTITVHDNRGLDREIHYHRATAEKNSDIRITRHHQYDVLGHLHQSIDP 60

Qy 61 RLYEAKQNTNTIKNFPLWOYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120
Db 61 RLYDAKQADNSKFNFWQYDVLGHALRTESVDAGRTVALNDIEGRPVMTWNAIVGRQTR 120

Qy 61 RLYEAKQNTNTIKNFPLWOYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120
Db 61 RLYDAKQADNSKFNFWQYDVLGHALRTESVDAGRTVALNDIEGRPVMTWNAIVGRQTR 120

Qy 121 QYETSSIPGRLLSVAEQTPPEKTSRITRILWAGNTAEKDNLAGOCVRHYDTAGVTRL 180
Db 121 RYEGNTIPGRLLSVSEQVFDQETSQVTERFIWENTTAEEKNLSGLCIRHYDTAGVTRL 180

Qy 121 QYETSSIPGRLLSVAEQTPPEKTSRITRILWAGNTAEKDNLAGOCVRHYDTAGVTRL 180
Db 121 RYEGNTIPGRLLSVSEQVFDQETSQVTERFIWENTTAEEKNLSGLCIRHYDTAGVTRL 180

Qy 181 ELSLTGTVLVSQSSQLLIDTQEAANWTGNETVWQNMMLADDIYTLTSTFDATGALLTQDA 240
Db 181 MSQSLAGAILSQSHQWLAEGQEAQNSGDETVWQRMMLASEVYTTQSTTNAIGALLTQDA 240

Qy 659 LNLRYMRVNNPVTLLDPDGLMPTI-----AER-----IAALKK- 691
```

```
Db 661 LNLRYMRVNNPISYHDNNGLHPNSNNNGLNERNHPLYYLDLAGVYVNVNDKEAILSTLKRS 720
Qy 692 -----NKVTDAPSAPNATNVAINTRPPVAPKPSLPKASTSQPTTHPIGAANIKPTT 744
Db 721 FAATSWGNKVSDDKTTQTTILSAEYGNLKHKKQMEYNNKLNKLNSTSPDGKFKYAKTKLODQA 780
Qy 745 SGSSIVAPLSPGVGNKSTSEISLPESAQSSSSTSTTNLOKKSFTLYRADNRSPREMOSKF 804
Db 781 AHAGIARNLHPISDNE-----KEGWTFTNYKDFVNLKOKESLGGKGLPGVELLPDN 833
Qy 805 PEGFK 809
Db 834 PEHYK 838

RESULT 8
ID Q7MZV7 PHOLL PRELIMINARY; PRT; 1043 AA.
AC Q7MZV7;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC1.
GN Name=tccC1; OrderedLocusNames=plu4167;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbr886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Tachout S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.";
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; BX571873; CAE16539.1; -; Genomic_DNA.
DR Photolista; plu4167; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0046677; P:response to antibiotic; IEA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000408; Reg_chromatophore.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1043 AA; 111877 MW; C3B871FAE134F53D CRC64;
```

```
Query Match 54.8%; Score 2744; DB 2; Length 1043;
Best Local Similarity 74.7%; Pred. No. 3.8e-138;
Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3;

Qy 1 MKNIDPKLYQKTPVSVVYDNRGLIIRNIDFHRITANGDPDTRITRHHQYDIGHLNQSIDP 60
Db 1 MSTPDTALYTQTPVSVVLDNRGLSIRDIGFHRVVGIDIDTRVTRHHQYDARGYLHNSIDP 60

Qy 61 RLYEAKQNTNTIKNFPLWOYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120
Db 61 RLYDAKQADNSKFNFWQYDVLGHALRTESVDAGRTVALNDIEGRPVMTWNAIVGRQTR 120

Qy 121 QYETSSIPGRLLSVAEQTPPEKTSRITRILWAGNTAEKDNLAGOCVRHYDTAGVTRL 180
Db 121 RYEGNTIPGRLLSVSEQVFDQETSQVTERFIWENTTAEEKNLSGLCIRHYDTAGVTRL 180

Qy 181 ELSLTGTVLVSQSSQLLIDTQEAANWTGNETVWQNMMLADDIYTLTSTFDATGALLTQDA 240
Db 181 MSQSLAGAILSQSHQWLAEGQEAQNSGDETVWQRMMLASEVYTTQSTTNAIGALLTQDA 240
```

Qy	241	KGNTORLAYDVAGQINGSWLTLKGQTEQVIKSLTYSAAQKGLREEHGNDVITREYSYEP	300
Db	241	KGNTORLAYDIAGQLKGSWLTIVKQREQQIVKLSWSAAGHKGLREEHGNGVITEYSYEP	300
Qy	301	TORLIGIKTR-----PSDTKVLQDLRYEYDVPGVNVISIRNDAEATFRWHNOKVMBENTYT	356
Db	301	TORLWGLITRRABESQSEARVLQDLRYEYDVPGVNVISIRNDAEATFRWNRQKVEPNRYV	360
Qy	357	YDSL YQLISATGREMANIGQOSHQPSPAL--PSDNNTVTNTYTRVTVYDRGGLNKTKIOHS	414
Db	361	YDSL YQLMSATGREMANIGQSQNLQPSAMPVPADNSITNTYQRTYSYDRGGLNVQIRHS	420
Qy	415	SPATONNTYNTIVSNRSRAVLSTLTEDPADQVDALFDAGGHQNTLI SQGNLNNWNRGEL	474
Db	421	SPATONSYTEMTVSSRSNRVTLSLTDTPTQVDGLFDAGGHOKTLIPQGNLDNNMRGEL	480
Qy	475	QOVTLVKDKGANDREWRYSGDGRMLKINEQOASNNACTQRTYTLPNLELRITQNST	534
Db	481	QRVTVFSREN--SSDSEWRYSSDGRVLKRVSEQQTGNSTQVRYTYLPGLERLITGVAD	538
Qy	535	ATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSOLELDSGQIIS	594
Db	539	KTONQLQVITVGEAGRAQVRVLHWSGKPADIDNNQVRYSYDNLIGSSQIELDSGQIILS	598
Qy	595	EEETYPYGGTALWAARNQTEASYKTI RYSGKERDATGLYYGYRYQYQPMIGRWLSSDPAG	654
Db	599	QBEETYPYGGTAVAAKNQTEASYKFI RYSGKERDATGLYYGYRYQYQPMVGRWLSADPAG	658
Qy	655	TIDGLNLYRMVNRNPVTLTDPGLMPT	681
Db	659	TVDGLNLYRMVNRNPITLTDRDGLAPS	685

RESULT 9  
OB85157 PHOLU  
ID OB85157 PHOLU PRELIMINARY; PRT; 1043 AA.  
AC OB85157;  
DT 01-NOV-1998 (TREMBLrel. 08, Created)  
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
DT 01-FEB-2005 (TREMBLrel. 29, Last annotation update)  
DE Insecticidal toxin complex protein Tccc (Toxin complex protein).  
GN Name:tcctc;  
OS Photorhabdus luminescens (Xenorhabdus luminescens).  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Photorhabdus.  
OX NCBI\_Taxid=239486;  
[1]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=W-14;  
RC MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;  
RX Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,  
RA Bharti R., firrench-Constant R.H.;  
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens.";  
RL Science 280:2129-2132(1998).  
[2]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=W-14;  
RC Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,  
RA firrench-Constant R.H.;  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RP STRAIN=W14;  
RC MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;  
RX Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,  
RA firrench-Constant R.H.;  
RL "The tc genes of Photorhabdus: a growing family.";  
RL Trends Microbiol. 9:185-191(2001).  
DR EMBL; AF047028; AAC38630.1; -; Genomic DNA.  
DR EMBL; AF346499; AAL18473.1; -; Genomic DNA.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:0046677; P:response to antibiotic; IEA.

DR	InterPro; IPR002345; Lipocalin.
DR	InterPro; IPR006530; YD.
DR	TIGRFAMs; TIGR01643; YD repeat 2x; 1.
DR	PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
SQ	SEQUENCE 1043 AA; 111694 MW; E467533FF34C548D CRC64;
	Query Match 54.78; Score 2740; DB 2; Length 1043;
	Best Local Similarity 74.77; Pred. No. 6.2e-138;
	Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3
Qy	1 MKNIDPKLYOKTPTVSVYDNRLGILIRINIDFHTTTANGDPDTRITRHQYDIHGLNQSIDP 60
Dd	1 MSPSETTLYTQTPTVSVDLNRGLSIRDIGFHRIIVGGDTDFRVTRHQYDARGHLNYSIDP 60
Qy	61 RLYEAKQNTNYIKENFWLDLTGNPLCTESIDAGRTVTLLNDIEGRPLLTVATATGVIOQR 120
Dd	61 RLYDAKQADNSKFENFWMQHDLAGHALRTSDVAGRITVALNDIEGRSYMWNATGVROQR 120
Qy	121 QYETSSLPGRLLSVAEQTPPEKTSRITERLLIWAGTEAEKDHNLAGQCVRHYDTAGVTRL 180
Dd	121 RYEGNTLPGRLLSVSEQVFQESAKVTERFTWAGNTTSEKEYNLSEL CIRHYDTAGVTRL 180
Qy	181 ESLSLTGTVLSSQSLLDITQEANWTGDNETVQNMLADDIYTTLTSTFDATGALLTQTDA 240
Dd	181 MSQS LAGAML SQHOLLAEQGANWSGDDETVMQCLASVYTTQSTTNALGALLTQTDA 240
Qy	241 KGMIORLAYDVAGOLNGSWLTLKQOTQOVIIKS LTYSNAAGOKLREEHGNDVITEYSYPE 300
Dd	241 KGMIORLAYDIAGLGKSLTVKGOSQOVIKSLWSAAGHKHREEHNGVVITEYSYPE 300
Qy	301 TORLIGHKTRR --- PSDTKVLQDLRYEYDPGVNVISIRINDAEATRFWHNKVKMPENTYT 356
Dd	301 TORLIGITTRAEGSQSGARVLQDLRYKYDPGVNVI SHINDAEATRFRWNKVFEENRYV 360
Qy	357 YDSLYQLISATGREMANIQGOHQFPSPAL -- PSDNNTYNTYTRYTYTDRGNLTKIOHS 414
Dd	361 YDSLYQLMSATGREMANIQGOSONLPSPVIPVPPTDDSYTYNLTYYTYTDRGNLVQIRHS 420
Qy	415 SPATQNNYTNITVSNRSRAVLSTLTEDPAQVDALPDAGHQNTLISGQNLNMNTRGEL 474
Dd	421 SPATQNSYTTDITVSSRSRAVLSTLTDPTRVDALPDSGHQXKLIPQGLDNWIRGEL 480
Qy	475 QQVTLVKRDKGANDREWYRSGDGRMLKINEQQASNNACTQRTVYLPNLIELRUTQNST 534
Dd	481 QRVTTPVSREN - SSDSEWYRYS DCMELLKVSEQQTGNS TQVRVYLPGLLELR TTGVAD 538
Qy	535 ATTDEDLOVITVGEAGRAQVRVLHWESCKPEDIDNNQLRYSVDNLIGSSOLELDSEGQIIS 594
Dd	539 KTTFEDLOVITVGEAGRAQVRVLHWESCKPTDIDNNQVRYSDNLIGSSOLELDSEGQIIS 598
Qy	595 EEERYYPGGTALWAARNQTESYKTIRYSGKERDATGLYYGYRYQPWGIRWLSSDSPAG 654
Dd	599 QEERYYPGGTAIWAARNQTESYKFIRYSGKERDATGLYYGYRYQPWPVGWRMLSADPAG 658
Qy	655 TIDGLNLYRMVRNPNVTLLDPDGLMPT 681
Dd	659 TVDGLNLYRMVRNPNITLTDHGLAPS 685

RESULT	10
ID	Q7N7Z1_PHOLL
ID	Q7N7Z1_PHOLL PRELIMINARY; PRT; 915 AA.
AC	Q7N7Z1;
DT	01-MAR-2004 (TReMBLrel. 26, Created)
DT	01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT	01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE	Insecticidal toxin complex protein TccC2.
DE	Name=tccC2; OrderedLocusNames=plu0960;
GN	Photorhabdus luminescens (subsp. laumondii).
OS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC	Enterobacteriaceae; Photorhabdus.
OX	NCBI_TaxID=141679;
RN	[1]



Db	293	ETQELIGIATRRPSDAKVLQDLRYQYDPGVGNVINIRNDAAETRFWRNQKVPENSYYTDS	352
Qy	360	LYQLISATGREMANIGQOSHOPPSPALPSDNNYTNVTRTYTYDRGGLTKIOHSSPATQ	419
Db	353	LYQLISATGREMANIGQONQOLPSPALPSDNNYTNVTRTSYSDHSGNLTKQIRUSSPATQ	412
Qy	420	NNYTTNITVSNRSRAVLSTLTEDPAQVDALFDAGGHQNTLISGNLNNWTRGELQQVTL	479
Db	413	NNYTVAITLSNRKNKGVLSLTITDPNQVDTLFDAGGHQTSLLPGQTLLTWTRGELKQV--	470
Qy	480	VKRDKGANDREWRYSGDGRMLKINFAQASNAQTORVTVYLPNLELRLTONSTATTED	539
Db	471	---NNGPGN--EWRYDSNGMRQLKVSQPTQNTTQQQRVLYLPGLLELRTTQSNATTTEE	525
Qy	540	LOVITVGEAGRAQVRVLHWESGKPEDIDNNOLRYSYDNLIGSSQLELDSBGQIISEEYY	599
Db	526	LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELDSBGQIISEEYY	585
Qy	600	PYGCTALWAARNQTEASYKTIYSGKERDATGLYYYGYRYQPMWGRWLSDDPACTIDGL	659
Db	586	PFGCTALWAANSQTEASYKTIYSGKERDATGLYYYGYRYQPMWAGRWLSADPACTIDGL	645
Qy	660	NLRYMRVNNPVTLDPDGLMPTIAERIAALKKNKVT	695
Db	646	NLYRMVNNPVSLQDENGAP-----EKGKVT	672

RESULT 12

Q7MZ20\_PHOLL

ID ID Q7MZ20\_PHOLL PRELIMINARY; PRT; 936 AA.

AC Q7MZ20;

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Insecticidal toxin complex protein Tccc7.

GN Name=tccc7; OrderedLocusNames=plu4488;

OS Photobacterium luminescens (subsp. laumondii).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Photobacterium.

OX NCBI\_Taxid=141679;

ON [1]

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=TT01;

RC MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt.886;

RX Tduchaut E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,

RA Draud A.S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,

RA Dassa E., Derosé R., Derzelle S., Freysinet G., Gaudriault S.,

RA Médigue C., Lanois A., Powell K., Siguiet P., Vincent R., Wingate V.,

RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;

RT "The genome sequence of the entomopathogenic bacterium Photobacterium luminescens";

RT Nat. Biotechnol. 21:1307-1313 (2003).

RL EMBL; BX571874; CAEI6860.1; -; Genomic\_DNA.

DR PhotolList; plu4488; -.

DR InterPro; IPR002345; Lipocalin.

DR TIGRfam; TIGR01643; YD repeat 2x; 2.

DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.

DR COMPLETE proteome.

SO SEQUENCE 936 AA; 103938 MW; 1B8B43F720798ECD CRC64;

	Query Match	47.7%	Score 2386;	DB 2;	Length 936;
	Best Local Similarity	52.0%	Pred. NO. 4.7e-119;		
	Matches	516;	Conservative 116;	Mismatches 227;	Indels 134; Gaps 23
Qy	10	QKTPSVYDNRGLIIRNIDFHTTAN-GDPDTTRITHQYDIHGHNLNOSIDPRLYEAKQT	68		
	:	:::::	:::::	:::::	:::::
Db	10	RNTFSVSVRDNRGLNIHTLRYLTQADETNSNELTRYQNTHGLQVKSTDPRRYK----	65		
Qy	69	NNTIKPFLWQDYLTGNPLCTYESIDAGRTVTLNDIEGRPLITVTATGVITRQHYETSSLP	128		
Db	66	-NOSGSNFTIFSLAGNTLRRESIDAGRTITLNDIEGRPVLITAINAIGAROTHHVEGNTLP	124		

Qy	129	GRLLSVABQTPPEEKTSRITERLIIWAGNTEAEKDHNLACQCVRHYYDTAGVTRLESLSLCT	188
Db	125	GRLLAVTELI--QRNEKTERLIIWANNTDAEKQNLAGQCIRHYDPAGLVQLESLSLGS	182
Qy	189	VLSSQLLIDTQBANWTDGNETVWONMLADDIYFTLSTFDATGALLTQTDAGKNIQRLA	248
Db	183	VLSSQRQLIADQADWHRGDDNSWRTKLNGNIPTTQHTDAIGALLTQIDAKGNQRLA	242
Qy	249	YDVAGQLNGSWLTLKGQTEQVIIKSLTVYSAAGQKLREEHGNDVITEYSEYPTQRLIGIK	308
Db	243	YDVAGQLKGSWTLKGQAEKVIQSIITWSAAGQKLREEHGNGVITEYTEYPTQRLINIT	302
Qy	309	TRRPSD--TKVLQDLRYEYDVPGNVISIRNDAEATFRWQKVPENTVYDLSLYOLISAT	367
Db	303	TRTRDSTKPLQDLRYEYDVPGNVINIRNDAEATFRWQKVLPENAISYDLSLYOLIQAT	362
Qy	368	GREMANIGQQSHQPPS--PALPDSNNTYNTTRYTYDRGGLNLTQIHSPPATONNTYN	425
Db	363	GREMANIGQGSQLPPLITPLPTDDNTYNTIRTYDYDSGNLTQIHSAPASNNNTYN	422
Qy	426	ITVNSRNRVLSLTITEDPAQVDALFDAGGHQNTLISGONLNNWTRFGLQOVLTKRDKG	485
Db	423	ITISNRNRRGVLSLTINDPNLIVDTFFDAGGHQTSLFSGOSLNNWTRFGLQOQV----	477
Qy	486	ANDREWRYSGDGRMLKINEQASNNNAQQRVTVLPNLELRLTQNSTATTEDLQVITV	545
Db	478	QNTAREWHYDSDGMRLLKINEQQTSTNTQOQVTVLPGLHLHTTQSGTNITEDLQVITV	537
Qy	546	GEAGRAQVRVLHWESGKPEDIDNNQLRYSYNLIGSSOLELDSBQOIISEEYYPYGGTA	605
Db	538	RQAGKAQVRVLHWBKGOPTGINNDQVRSYDNLTHSSSELELDMHGEMISWEEYYPYGGTA	597
Qy	606	LWAARNOTEASYKTIIRYSGKERDATGLYYGVRYQPMIGRWLSSDPAGTIDGLNLYRMV	665
Db	598	VWAARNOIEAGYKTIIRYSGKERDATGLYYGTRYQPMAGRWLSADPATVQDGLNLYRMV	657
Qy	666	RNNPVTLLDPDGLM--PTIAERIAAL-----KKNKVTDSPASPANATNVAINIRPEV	715
Db	658	RNNPMTGIDEDGRMFKTVATGALGIGGMAYELKYKQOQVE-----	698
Qy	716	APKPSLPKASTSSOPHTHPIGAANI--KPTTSGSSIVAPLSPGVNKSTSEISLPESAQS--	773
Db	699	--KKPMP-----SPVSSEYGDQOVSKITQKASALKANYDPMQMAHNITGRKEAEHLA	750
Qy	774	SSSTTSTNLQKSTFLYRADNRSPEEMQSKPEGFKAWTPLDTKMAROFASI-FIQOKDT	832
Db	751	QGOVPGQAIKEGASLVAA-----AGEFAAGNVGELNK	784
Qy	833	SNLPK-----ETVKNISTWGA-----KPKLKDLSNYIKVTKDKSTWVWST	872
Db	785	TTLTKAAGTDAINNFSUTLDGIKVKFKAGASAATVSPKELBELQ---KATDDLKEVATDT	841
Qy	873	AINTAAGOSSGA-----PLHKIDMDLVEFA--IDQ-----OKNLNPLPEGRTKN	914
Db	842	LITGASIGATVGATLDTAAAVVPHPVAKVALKGLSLAWKVTGIVHTAELSLAE--KHK	900
Qy	915	MVPSLL-----LDTPOIETSSIIA--LNHG	937
Db	901	LVSNELGOELKREVTANQIRRGITILAKVRSHG	933

RESULT 13	
Q6QIY7 KENNE	
ID Q6QIY7 KENNE PRELIMINARY;	PRT; 1016 AA.
AC Q6QIY7	
DT 05-JUL-2004	(TRMBLrel. 27, Created)
DT 05-JUL-2004	(TRMBLrel. 27, Last sequence update)
DT 05-JUL-2004	(TRMBLrel. 27, Last annotation update)
DE TccC1/XptB1 protein (Fragment).	
GN Name=tccC1;	
OS Xenorhabdus nematophilus.	
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriaceae; Yersiniaceae; Yersinia	







```

Db 353 TYIVDSLYQLVSATGREMANAGQGNLPSATAPLPTDSSATYNTYRVRDRGNLTOM 412
Qy 412 QHSPATONNTYNTITVSNRNRVLSTLTEDPAQVDALFDAGGHONTLISQNLNMNTR 471
Db 413 RHPATNNNTYNTITVSDRNRVLSTLAEPVSDVLMFSGGHOKHLQPGQALVWTPR 472
Qy 472 GELQVTLVGRDKGANDREWRYSGDGRMLKINEQOASNAQRTVYLPNLELRTQ 531
Db 473 GELQVTPVVRDGA--DSESRYDAGSQRIIKTGTROTGNVQRTQVVVLPGLRLMA 531
Qy 532 NSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSOLELDSGQ 591
Db 532 NGVTEKESLOVITVGEAGRAQVRVLHWEIGKPDDESDVSRYSDNLVGSOLELDRGY 591
Qy 592 IISSEYYPYGGTALWAARNQTEASYKTIYSGKRDATGLYYGYRYYPQWIGRLSSD 651
Db 592 LISEEFYPYGGTAVLTARSEVADYKTIYSGKRDATGLDYGYRYYPQWAGRWLSTD 651
Qy 652 PAGTIDGLNLYRMVRNPNVTLDDPGLMPT-----IAERIAALKK 691
Db 652 PAGTVDGLNLFWRVNRNPNVTLFDSNGRISTGQEARRLVGEAFVHPLHMPVFERISVERK 710

RESULT 15
Q66A99 YERPS
ID Q66A99 YERPS PRELIMINARY; PRT; 994 AA.
AC Q66A99
DT 25-OCT-2004 (TREMREL. 28, Created)
DT 25-OCT-2004 (TREMREL. 28, Last sequence update)
DE Putative insecticidal toxin complex.
GN OrderedLocusNames=YPTB2233;
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP32953 / Serotype I;
EX PubMed=15358858; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,
RA Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francois V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RT "Insights into the evolution of Yersinia pestis through whole-genome
comparison with Yersinia pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
DR EMBL; BX936398; CAH21471.1; -; Genomic_DNA.
DR InterPro; IPR005165; Anthrax_toxA.
DR InterPro; IPR006530; YD.
DR Pfam; PF03497; Anthrax_toxA; 1.
DR Pfam; PF05593; RRS_repeat; 4.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW Complete proteome.
SQ SEQUENCE 994 AA; 109260 MW; 76C8CE064806F46F CRC64;

```

```

Query Match 43.4%; Score 2173; DB 2; Length 994;
Best Local Similarity 53.3%; Pred. No. 1.3e-107;
Matches 451; Conservative 105; Mismatches 208; Indels 82; Gaps 14;

Qy 8 LYQKPTVSVVDNRGLIIRNDFHRTTANGD-PDTRITRHOYDIGHNLNQSIDPRLYEAK 66
Db 5 LFSKTPSVTVLDNRGLTRDIAHRRPDPVTSERITHQYDARGFLTQSADPRLHGAG 64

Qy 67 QTNNTIKPNFLWQYDLGNPLCTESIDAGRVTTLNDIEGRPLLVTT-----AT 114
Db 65 LM-----NFSYLTDLTGRIILRTQAGDNGTTVSLNDATGRPPISASNISSDSDGTEDRGQ 118

Qy 115 GVITQROYETSLPGRLLSVAEQTEPEKTSRTERLIWAGNTEAKOHNLAGOCVRYHYDT 174
Db 119 AMTRTWQYEASLPGRLVSVTEQV-TGKATRITERFVYAANTDAEKSINLAGACVSHYDT 177

```

```

Qy 175 AGVTRLSLSLTGTVLSOSSQLLIDTQE-----ANWGTGNETVWQMLADDIYTTLTSTFPA 230
Db 178 AGLVQPSIALTGVPLSVTRLMKSADNPDAVADWQGADASAWNDQLDGETHTTLTTTADA 237
Qy 231 TGALLTOTDAKGNITORLAYDVAGOLNGSWLTKGTEQVVIKSLTYSAAQOKLREHGND 290
Db 238 TGAULTTTDAKGNIQRMAYDVAGLLSGSWLTKGTEQVIVKSLTYSAAQOKLREHGNG 297
Qy 291 VITEYSYBETQRLIGIKTRRP-----SDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHN 346
Db 298 VVTTYEYEPETQRLVGIKTERPAGHASGAKVLQDLRYEYDPVGNVLKVTNDAEETRFWRN 357
Qy 347 QKVPENTYTVDSLYQLISATGREMANIGQSHQPPSPA--LPSDNNTYNTYTRYTYDR 404
Db 358 QKVPENTYTVDSLYQLVSATGREMANAGQSCSLPSTTVPLPADSSAYTRYSTRYTYDE 417
Qy 405 GGNLTKLOHSPATONNTYNTITVSNRNRVLSTLTEDPAQVDALFDAGGHONTLISQ 464
Db 418 AGNLTQIRHNAPATNNSYTTIKITVSDRNRGVLSTLTENAADVADALFTAGGQOTQLQPGQ 477
Qy 465 NLNWNTRGELQOQVTLVKRDKGANDREWRYSGDGRMLKINEQOASNAQRTVYLPN 524
Db 478 HLIWTARNELLKVTVPVRED--GSTDDSESRYDAASQRIKVSROKTNNTSMQTRVLYLPG 536
Qy 525 LELRLTONSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSQL 584
Db 537 LELRSKSGDTETEGLOVITVGEAGRAQVRVLHWSGKPEDIDTDDQIRYSYDNLGSSCL 596
Qy 585 ELDSEGOIISSEYYPYGGTALWAARNQTEASYKTIYSGKRDATGLYYGYRYYPQWI 644
Db 597 ELGGDGNIIISAEYYPYGGTAVWVRRAVEADYKTIYSGKRDATGLYYGYRYYPQWA 656
Qy 645 GRWLSSDPAGTIDGLNLYRMVRNPNVTLDDPGLMPTIAERIAALKKVKVTDSPAPANA 704
Db 657 GRWLSADPAGMVDGLNLFWRMARNPNVAFIDRNLNSELAYS-QAFKRT-----ANK 706
Qy 705 TNVAINRPPVAPKPSLPKASTSSQPTHTPIGAANIKP--TTSGSSIVAPLSPVG----- 757
Db 707 YNVIIGVRAP-----NPLGETLLKEGFPFSKNPHMKAKSSPTGTAGF 748
Qy 758 ---NKSTSEISLPESAQSSSSSTTSTNLQKSFITYRADNRSFE-----EMQ 801
Db 749 IAEPTIYKVSFSAKKQORASIDKAKALGSESIDLFISKSRINELIDTGNLSLGENRYS 808
Qy 802 SKFPEG 807
Db 809 AKYPVG 814

```

Search completed: February 16, 2006, 21:43:00  
Job time : 133.785 secs

**THIS PAGE BLANK (USPTO)**

---

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2537.5	50.7	915	6	US-10-647-956A-6	Sequence 6, Appli
2	183	3.7	2828	7	US-11-080-991-54	Sequence 54, Appl
3	183	3.7	2828	7	US-11-186-284-49	Sequence 49, Appl
4	175	3.5	2725	7	US-11-113-424-52	Sequence 52, Appl
5	174.5	3.5	1417	7	US-11-052-554A-8	Sequence 8, Appli
6	171.5	3.4	1345	7	US-11-052-554A-282	Sequence 282, App
7	164.5	3.3	5024	6	US-10-793-626-2964	Sequence 2964, Ap
8	160.5	3.2	1579	7	US-11-052-554A-9	Sequence 9, Appli
9	156	3.1	2314	7	US-11-013-759-11	Sequence 11, Appl
10	154.5	3.1	1468	6	US-10-467-657-1088	Sequence 1088, Ap
11	153.5	3.1	2769	7	US-11-113-424-14	Sequence 14, Appl
12	153	3.1	1045	7	US-11-113-424-54	Sequence 54, Appl
13	153	3.1	1094	6	US-10-821-234-1097	Sequence 1097, Ap
14	153	3.1	2376	7	US-11-096-051-4	Sequence 4, Appli
15	153	3.1	2715	7	US-11-096-051-2	Sequence 2, Appli
16	153	3.1	2715	7	US-11-113-424-51	Sequence 51, Appl
17	153	3.1	2721	7	US-11-096-051-10	Sequence 10, Appl
18	153	3.1	2725	7	US-11-096-051-8	Sequence 8, Appli
19	152.5	3.0	1562	7	US-11-052-554A-211	Sequence 211, App
20	152.5	3.0	2515	7	US-11-113-424-53	Sequence 53, Appl
21	151.5	3.0	2053	7	US-11-013-759-9	Sequence 9, Appli
22	151	3.0	5251	7	US-11-052-554A-281	Sequence 281, App
23	150.5	3.0	1992	7	US-11-013-759-3	Sequence 3, Appli
24	150.5	3.0	1992	7	US-11-013-759-13	Sequence 13, Appl
25	150.5	3.0	2047	7	US-11-013-759-4	Sequence 4, Appli

```
Qy 300 ETORLIGKTRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFHWNKMPENTYTYDS 359
Db 293 ETORLIGITTRPSDAKVLQDLRYQYDPVGNVINIRNDAEATRFWRNKVAPENSYTYDS 352
Qy 360 LYQLISATGRMANIGQOSHOPFPALPSDNNVTYTRTYDRGGLTKIQRHSSPATQ 419
Db 353 LYQLISATGRMANIGQONQLPSALPSDNNVTYTRSYDHSNGLTQIRHSSPATQ 412
Qy 420 NNYTNTITVNSRNLAVLSTLTEDPAQVDALFDAGGHQNTLISGNLWNRGELQVTL 479
Db 413 NNYTVAITLNSRNRGLVLTLPDNPQVDTLFDAGGHQNTLISGNLWNRGELQVTL 470
Qy 480 VKRDKGANDREWYRSGDGRMLKINEQOASNAQOTVYLPNLHLRTONSTATED 539
Db 471 ---NNGPGN---EWYRDSNGRQLKVSQEQPTQOORVYLPGLRLTQSNATTEE 525
Qy 540 LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIIEBEYY 599
Db 526 LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELDNQOIIIEBEYY 585
Qy 600 PYGCTALWAARNQTEASYKTIYRSGKERDATGLYYYGYRYQPWIGRWLSDPDAGTIDGL 659
Db 586 PFGCTALWAANSQTEASYKTIYRSGKERDATGLYYYGYRYQPWAGRWLSADPAGTIDGL 645
Qy 660 NLYRMVRNPNVTLDPGLMPTIAERIAALKKVKVT 695
Db 646 NLYRMVRNPNVSLQDENG LAP-----EKGYT 672

RESULT 2
US-11-080-991-54
; Sequence 54, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-080-991-54

Query Match 3.7%; Score 183; DB 7; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.0017;
Matches 165; Conservative 108; Mismatches 276; Indels 302; Gaps 38;

Qy 307 IKTRRPSDTKVLQ---DLRYEYDPV-----GNVISI-RNDAEATRFHWNQ 347
Db 542 IKSMPSDSGLYQCTAQVRDEMDRMVYRVLVQSPSTQPAEKDTVTIGKNPGESVTLPCNA 601
Qy 348 KVMPEYTYDLSLQLISATGRMANIGQOSHQPSP-----ALPSDNNVTYTRTYTYD 403
Db 602 LAIPEALHSM-----ILPNRRIINDLANTSHVYMLPGLTSLPKVQVSDSGYTRCVAVN 655
Qy 404 RGGNLTQIOWSSPATONNTYNTIVS-----NRSNRAVLSTLTEDPAQVDALFD 452
Db 656 QQG-----ADHFTVITVTKGSLPSKGRRRPGAKALSRVREDIVE-----D 698
Qy 453 AGGHQNTLISQNLWNRGEL-----QQVTLVKRDKGANDREWYRSGDGRMLKI--- 505
Db 699 EGG-----SGMGDEENTSRLLHPKQDQEVFLTKDDAINGDKK-----AKKGRKJLWK 748
Qy 506 --NEQOASNAQOTQV-----T 520
```

```
Db 749 HSEKEPTNVAEGRVFESERRINMANKQINPERWADILAKVRGNLPGKTEVPPLIKTT 808
Qy 521 YLPNLELRLTON-----STATTEDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYD 576
Db 809 SPPSLSELEVTPPPFPAVSPSASPVQTVTSAESSADVPLLEE-----E 852
Qy 577 NLIG-----SQLELDSEGOIIESEYYPYGGTALWAARNQTEASYKTIYRSGKERDATG 631
Db 853 HVLGTIISASGMLHNNHNGVILPEV-----TSTPLEEVVDDLSKTEETITSTEGDLKG 907
Qy 632 -----LYYGYRYQY-----WIGRWLSDPDAGT-----IDGL 659
Db 908 TAAPTLISEPEYSPPTLHLLDTVYKETHETATEGWSAADVGSSPEPTSEYEPPLDAV 967
Qy 660 NLYRMVRNPNVTLDPGLMPTIAERIAALKKVKVTDSPSP-----ANATVAINIRPPVA 716
Db 968 SL---ABSEPMQVFPD-LETQSQPDEDKMKEDTFAHLLTPTTIWVNDSSSTQLFEDSTI 1023
Qy 717 PKPSLP-----KASTSSQPT-----THPIGAANI---KPTTSGSS 748
Db 1024 GEPGVGQSHLQGLTDNIHLVKSLSLTDLLIKKMKEMSQTLOGGNMLEGDPHRSRS 1083
Qy 749 -----IVAPLSPVG-----NKSTSEISLPSAQSSSSSTTS 779
Db 1084 ESEGOESKSIPLDSTLIGINSSMSPVKKPAETTVGTLDDKDTTVTTTPRKVAPSSTMS 1143
Qy 780 TNLQKKSFTLYRADNRSFEBEQKPFEGFKAWPLDTPKMARQFASIFIGKQDNTNLKET 839
Db 1144 THPSR-----RRPNRRRLRPKNKFRHRKQTPP-----TTFAPSET 1179
Qy 840 VKNISTWGAPEKLKOLSNYIKYTKDKSTVWVSTAINT-----EAGGSSGAP----- 886
Db 1180 FSTQPT--QAPDIK-ISSQVE-SSLVPTAWVDNTVNTPKOLEMKNAPETS KGPTRKHG 1235
Qy 887 ----LHKIDMDLYEFAIDGQKLNPLPEGRTKNMVPSLLLDTPQIET-----SSIIALNHGP 938
Db 1236 KRPKNKHYTSTVSSRASGSKSPSPENKHNIV-----TPSSETILLPTVSLKTEGP 1289
Qy 939 VNDABISFLTT 949
Db 1290 YD--SLDYMTT 1298

RESULT 3
US-11-186-284-49
; Sequence 49, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamathkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2ERNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 49
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-49

Query Match      3.7%; Score 183; DB 7; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.0017;
Matches 165; Conservative 108; Mismatches 276; Indels 302; Gaps 38;

QY 307 IKTRPSDVKVLQ---DLREYEDPV-----GNVISI-RNDAEATRFWNO 347
Db 542 IKSMPSDGLYQCTAQAQVDEMDRVMYRVVLSQSPQPAEKDTVTIGKNGPESVTLPCNA 601
QY 348 KMPENTTYDLYQLISATGREMANIGQSQHFPSP-----ALPSDNNTYTYTYTD 403
Db 602 LAIPEALSW-----ILPNRIINDLANTSHVYMLPNTLSIPKVOVSDSGYRCVAVN 655
QY 404 RGNLTKIOHSSPATONNTYTVS-----NRSNRAVLSTLTEDPAQVDALFD 452
Db 656 QQG-----ADHFTVGTITVTKGSLGSPKRGRRPGAKALSRVREDIVE-----D 698
QY 453 AGGHQNTLISQNLNNTNGEL-----QQVTLVKDKGANDREWVRYSGDGRMLKI-- 505
Db 699 EGG-----SGMDENTSRLLHPDKQEVFLTKDIDAINGDK-----AKGRRKLKLWK 748
QY 506 --NEQASNNACTQRV-----T 520
Db 749 HSEKEPTNVAAGRVRFSRRRINMANKQINPERWADILAKVRGNLPGKGTVPPLIKTT 808
QY 521 YLPNLELRLTON-----STATTEDLQVITVGEAGRAQVRVLHWSGKPEDIDNNQRYSD 576
Db 809 SPPSLSLVTPPPAVSPSPSASPVQTVTSAESSADVPLGEE-----E 852
QY 577 NLIG-----SSOLEDSQIISSEYYPYCGTALWAARNQTEASVKTIRYSGKERDATG 631
Db 853 HVIGTISASMGLEHNGHNVILVEPV-----TSTPLEVVDLSEKTEBITSTEGDLKG 907
QY 632 -----LYYGYRYYPQ-----WIGRWLSSDPAGT-----IDGL 659
Db 908 TAAPTLISEPYEPSPLHTLDVTYKPTHEETATEGMSAADVGSSPEPTSSSEYEPPLDAV 967
QY 660 NLYMRVNNPVTLLPDGLMPTIAERIAALKKQKVTDSAPSP-----ANATNVAINTRPPVA 716
Db 968 SL---AESEPMQYFPD-LETQSPQDEKMKEDTFAHLTPTTIWVNDSSSTSQLPEDSTI 1023
QY 717 PKPLP-----KASTSQPT-----THPIGAANI---KPTTSGSS 748
Db 1024 GEPGVGQSHLOGLTNTLHLVKSSUSTQDTLLIKGMKEMSQTLQGGNMLEGDPHRSRS 1083
QY 749 -----IVAPLSPVG-----NKSTSEISLPESAQSSSSSTTS 779
Db 1084 ESEGESKSIITLPSDTLGMSSMSPVKKPAETTVGTLLDKDTTITVTPRKVAPSSTWS 1143
QY 780 TNLQKKSFTLRADNRSPEEMOSKEPEGKAWTPLDTKMAQFASIFIGQKDTSNLPKET 839
Db 1144 THPSR-----RRPNGRRLRNKFRHRKQTPP-----TTPAPSET 1179
QY 840 VKNISTGAKPKLKDLSNVIKYTKDKSTVWYSTAINT-----EAGQSSGAP----- 886
Db 1180 FSTQPT--QAPDIK-ISSQVE--SSLVPTAWVDNTVNTPKQLEMEKNAEPTSGTPRKHG 1235
QY 887 ----LHKIDMDLYEPAIDQKLNPLPEGRTKMVPSSLTDPQIBT-----SSIIALNHGP 938
Db 1236 KRPKNHRYTPPSTVSSRASGSKSPSPENKHNIV-----TPSSETILLPRVSLKTEGP 1289
QY 939 VNDABISFLT 949
Db 1290 YD--SLDYMTT 1298

RESULT 4
US-11-113-424-52

```

```

; Sequence 52, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2725
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-52

Query Match      3.5%; Score 175; DB 7; Length 2725;
Best Local Similarity 22.3%; Pred. No. 0.0053;
Matches 173; Conservative 85; Mismatches 256; Indels 262; Gaps 44;

QY 18 YDNRLGIIRNIDFHRITANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNTTIKPFL 77
Db 1865 YDQSGKII-----SRTWADG---KIMSYTY-----LEKSVMLLLHSQRR-----YI 1902
QY 78 WOYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQ-----TRQYETSSLP 128
Db 1903 FEYD---QPDCLLS-----VTMPSWVRHSLSQTMLSVGYRNIYTPPSSTSFIDYSRD 1953
QY 129 GRLLSVASQTPPEKTSRITERLIWAGNTEAEKHNLACQVRHYDTAGVTRL--ESLSLT 186
Db 1954 GRLL-----QTLHLGTGR---RVLYKYTKQARLSEVL-----YDTTQVTLTYESSGV 1999
QY 187 GTVLSSQSQLID-----TOEANWTGD-----NETVQNMALADDIYTLTSTFDATG- 232
Db 2000 KTI-----HLMHDPICITRYRQTGLPIGRQIFRESEGLVNARFDYSY---NNFRVTSM 2051
QY 233 -ALLTQTDAKNIQRLAYDVAGQLN--GSWLTGKQTEQVIKSL-----TVSAAGOKLR 284
Db 2052 QAVINETPLPDLRYV-DVSGRTEQFGKFSVINDLQVITTTVMKHTKIPFANGQ--- 2107
QY 285 BEHGNVDITEYSYRPTQRLIGIKTRRPSDTKVLQDLRY---EYDPVGNV-----ISIRN 336
Db 2108 -----VIEQVE-----ILKAIAYWMTIQVDNVRMVICDIRVGV 2142
QY 337 DAEATRFWHNOKVMPENTTYDLYQLISATGREMANIGQSQHFPSPALPSDNNYTYNY 396
Db 2143 DANITRYFYE-----YDADGQLQTVSV-----NDKTCW 2170
QY 397 TRTTYDRGNLTQIOHSSPATONNTYTNITVSNRNRNAVSLTLEDPAQVDALDAGCH 456
Db 2171 --RYSYDLNGNINLLSHGKSA-----RLTFLRYDLRDRITRLGEIQYKMDS--DGFLR 2219
QY 457 QNTLISQNL--NWNTRGELQVTLVKRDKGANDREW---YRSGDGRMLKINKINQQASN 512
Db 2220 QR-----GNDIPEYNSNGLQ-----KAYNKASGWTVQYYDGLGRRV-----ASK 2260

```

```
QY 513 NAQTQ-----RVTYLPNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWE 559
D 2261 SSLGQHLQFFVADTNPVRVTHLYN---HTSSEITSLYYDLDQ-----GHLIAEMLS 2308
QY 560 SKPEDIDNNOLRYSYDNLIGSSQLELDSEGOITSEBEYYPYGGTALWAARNOTEASVKT 619
D 2309 SGE-----EYVACDN-TGTPLAVFSSRGQVKEILYTYGD-----IYHDTYDPFQV 2355
QY 620 I--RYSGKERDATGLYYGYRYQPWIGRW-----LSSDPAGTIDGLNLYRMV 665
D 2356 IIGPHGLYDFLTKLVHLGQRDYDVVAGRWTPPNHHIWKQLNLLPKP-----ENLYSFE 2409
QY 666 RNNPVTLDDPGLMPTTAERTAAALKKQVTDASAPANATVAINIRPPVAPKSL 721
D 2410 NNYPV-----GKIQDVA-----KYTTDIRSWLELFGQLHNVLPGFPPKPEL 2450

RESULT 5
US-11-052-554A-8
; Sequence 8, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 8
; LENGTH: 1417
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-8

Query Match 3.5%; Score 174.5; DB 7; Length 1417;
Best Local Similarity 19.3%; Pred. No. 0.0022;
Matches 209; Conservative 122; Mismatches 327; Indels 427; Gaps 52;

QY 107 PLLTWTATGVIQTROYTSSLPGRLLSVAEQTPBEK---TSRITERLIWAGNTE--AEKD 161
D 291 PLLTLSA-GHKQKSGENDTRFGLVNVYRIGEPFLAKQLDTSIRERRVLGSRDYDLVERN 349
QY 162 HNLAQCVRHVDYTAGVTEL-----EELSITGTVLSQSSQLLIDTQEAHW----- 205
D 350 NNI-----VLEYRKSEVIRIALPERIEGKGOTLSL-GLVSVKATHGLKNVQ---WEAPSL 401
QY 206 -----TGDNETVWQNMJL-----ADDIYTLSTFDATGALLTQTDAGNI-QRLAYDV- 251
D 402 LAEGKLTGQG-SQWQVTLPAVRPKONYAISA-----VAVDNKGNTSKRVQTEVV 452
QY 252 ---AG-QLNGSWLTLKGQTE-QVIKSLTYSAAGQKUREEHGNDVITEYSVEPTQRLIG 306
D 453 ITGAGMSADRTALTLDGQSRIQMLANGNEQKPLVLSLRDARG-----QPVTGMKDQ 503
QY 307 IKTRRPSTDKVLQDLRYEYDVGNVISIRNDAEATRFHWNOKVMPENTYTYDLSYQLISA 366
D 504 IKT-----ELTFKPAIGNIVT--RSLKATKS-QAKPTLGEFTETAGVQSFT 548
QY 367 TG-----REMANI-----GOOSHQFPSPA 385
D 549 TGTQSGRATITVSDGMSKVTYAEIRATMMDVANSTLSANPESGDVADGQQAATLTILTA 608
QY 386 LPSDNNVTYN-----YTRYTYDRCGNLT----- 409
D 609 VDSEGNPVTGASRLRFVPOPTNGVTGVAISEIKPGVYSAAVSVSTRAGNVVVRFAFSEQYQ 668
```

```
QY 410 -----KIOHSS-----PATQNNYTTNITVSNRSNRAVLSTLTEDPAQV 447
D 669 LGTTLOQTILKFVAGPLDAAHSITILNPKPVPVGGTVAITWTKDAYDNPVTSLTPEAPSLA 728
QY 448 DALFDAGGHQNTLISGQNLWNTRGE---LQQVTLVKRDKGANDRDREYRYSGDGRMLK 504
D 729 GA-----AABEGSTASG---WTNNGDGTWTAQITL-----GST-----AGELEVMPK 766
QY 505 INEQOASNNACTQRTVTLNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWSCKPE 564
D 767 LINGQAAANA--AKVTVVAD--ALSSNOSKVSVAED-----HYKAGEST 806
QY 565 DIDNNQLRYSYDNLIGSSQLELDSEGOIISBEYYPYGGTALWAARNOTEASYKTIRYSG 624
D 807 TV-----TLVAKDAHNAIS-----GLALSASLTGTASEGATV--SS 841
QY 625 KERDATGLYYGYRYQPWIGRWLSSDPAGTIDGLNLYRMVRNPNVTLDDPDGLMPTIAE 684
D 842 WTEKNGSY-----VATLTGKGTGELRVMPLF-----NGQPAATEAAQLTVIAG 886
QY 685 RIAALKKQKVTDS-APSPANATVAINIR-----PPVAPKESLP-----KASTSSQ----- 729
D 887 EMSSANSTLVADNKAFTVKTITELFTVKQDAYGNPVTGLKPDAPVFGAASTGSRPSAG 946
QY 730 -----PTTHPIGAA---NIKPTTSGSSIVAP---LSPVGNKSTSEI----- 764
D 947 NWTGKNGVYVSTLTLSAAGQLSVMPRVNGONAVAQPLVNLVAGDASKABIRDWTVKVN 1006
QY 765 -----SLPESAQSSSSSTSTN----- 781
D 1007 NQLANGQSANQITLTVVVDTYGNPLQGOEVLTLPGQVTSKGTGNTVTTNAAGKADIELMST 1066
QY 782 -----LOKSFILYRAD-----NRSFEEMOSKFPFGKAWTPLDTKMARQF 822
D 1067 VAGEHNISASVNGAOKTTVTVKFNADASTGOANLQVDAQAQVANGKDAFT--LTANVEDKN 1125
QY 823 ASIFIGQKDTSNLPKETVKNISTWGAQPKLKLDSNYIKYTKDKSTVWVSTAINTEAGGQS 882
D 1126 GNPVPGSLVTFNLP-----GVKPLTCD-----NVWV----- 1152
QY 883 SGAPLHKIDMDLYEFAIDGQKLNPLPEGRTKNNMVPSSLDDTPQIETSSIIALNHGPVND 942
D 1153 -----KAN--DEGKAEQLQVSVTAGTYEITAS---AGNSQPSNTQ 1187
QY 943 EISPL 947
D 1188 TITFV 1192

RESULT 6
US-11-052-554A-282
; Sequence 282, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 282
; LENGTH: 1345
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-282

Query Match 3.4%; Score 171.5; DB 7; Length 1345;
```

Best Local Similarity 19.4%; Pred. No. 0.0032;  
Matches 219; Conservative 140; Mismatches 367; Indels 401; Gaps 55;

QY 3 NIDPKLYQKTPVSVYDNRGLIIRNIDFHRHTTANGDPDTRITRHOYDIHGLNQSIDPRL 62  
DB 14 NNDTKVNIAPAS-----NAQVTLNIPAOVVVNNSDSVQLTATVKDPSNHPVAGI----- 65

QY 63 YEAKQTNNTIKPNFLQWYDLTGNPLCTESIDAGRTVTTLNDIEGRPLLTATGVIQTRQY 122  
DB 66 ----TVNFTMPQDVAAANFTLENNGIAITQANGEAHVTL---KGKAGTHVTATLSNNNT 118

QY 123 ETSSLPGLLSVAEQTPPEKTSRITERLIWAGTEAEKDHLAGOCVRYHDYTAGVTRLES 182  
DB 119 SDSQ---PVTFFVADKTSALVQLQISKNEI-TGN-----GV---DS 151

QY 183 LSLGTVLSSQSOLLIDTQEANWTGDNETVQNMMLADDIYTLSTFDATGALLTQTDAGK 242  
DB 152 ATLTAIVKQDF-----DNE-----VNNLPVTFST-ASSGLTLFPGESNT 189

QY 243 NIQRLAYDVAGOLGSMWTLKQGT--EQVILKSLTYSAAQOKLREEHGNVDITEYSYPE 300  
DB 190 NESGIAQ-----ATLAGVAFGEQTVTASLANNGAS-----DNK 222

QY 301 TORLIGIKTRRPSDKVLQDLRYEDPV-----GNVISIRNDAEATRFWNOQ 347  
DB 223 TVHFIG-----DTAAAKII--ELTPVPDSIIAGTPQNSSGSVIT-----ATVVDNNG 267

QY 348 KVPMENTYDLSYQLISATGREMANIGQSHQSPALPSDNN---TYTNITRTTYD 403  
DB 268 FPKGVTVNFTS-----NAAATMNGGO-----AVTNEQKATVITYN-TIS-SIE 312

QY 404 RGGNLTKIQHSPATQNNYTNITVSNRSRAVLSTLTEDPAQVDALFD---AGGHQNTL 460  
DB 313 SGARPDVTEASLENGSSTLSINVNADASTAHLTL-----QALFDTVSAGDTTLY 365

QY 461 ISGQNLNWNTRGELQOQVTL-VKRDIG-----ANDREWY-RYSGDGRMLKINEQ 508  
DB 366 IEVKD-NYCGNVPQOEVTLSPSPGVTGPNNAIYTNHNGNFYASFTAKAGVYQVAT 424

QY 509 QASNAQQRVTLPLNL---ELRL-----TONSTATT----- 537  
DB 425 LENGDSMQOVTYVFNVAEISLAASKDPVIANNDITLTATVADTEGNAIANSEVTF 484

QY 538 ---EDLQV-ITVGEAGR-----AQVRVLHWSGKPEIDNN----- 569  
DB 485 TLPEDEVANFTLGDGKVVDTEGKAVTLKGTAKAGHTVTASMAGGSEQLVWNFIADT 544

QY 570 -----OLRYSVDNLI-----LDSEGOIISBEY-----YPYGGT 604  
DB 545 LTAQVNLAVNTEDNFIANNVGMTRQLQATVTDGNGNPLANEAVTFLPADVSASFILGQGS 604

QY 605 ALWAARNQTEASYKTIIRSGKERDATGLYYGYRYQPWIGWLSDDPAGTID---GLNL 661  
DB 605 AITDINGKAEVT-----LSGTSKGYTPVTVSVNN 633

QY 662 YRMVRNPNFTLDPDGLMP-----TIAERIAALKKKNKVTDSAPSANATVAINI 711  
DB 634 YGVSTKQVTLIADAGTAKLASLTSVSVFVSTTEGATWTASVTDANGNVE--GIKYNE 691

QY 712 RPPVAPKPSLPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPGVNKSSTSEISLPESAQ 771  
DB 692 R-----GTSVTLSSTSVE--TDDRGFABELITSTEVGLKTVSASLADKPTVISRLNAK 744

QY 772 SSSSTSTTNLKKSFYLRADNRFRFEMQKPEG----- 807  
DB 745 ADINSATITSL-----IPEGQVNVADVAHVANDQFQNPILN 784

QY 808 ----FKAWTPLDTKMAROPASIFIGOKTSLNPKETV---KNISTWGAKPKLKOLSNYIK 860  
DB 785 BSVTFSAEPPEHMTISQNIVS-----TDTHTGIAEVTWTPERN-----GSVMKASLANGSS 835

QY 861 YTKD-----KSTVWVST---AINTAAGG-----QSSGAPLHKIDMDLYEFAIDGQKL 904  
DB 904

DB 836 YEKDLVVIDQKLTLSASSPLIGVNSPTGATLTATLTSANGTP-----VEGQVI 883

QY 905 N--ELPEGR-----KMWPSLLDTPQIETSSIIALNHGPV 939  
DB 884 NFSVTEPAGTSLGGKVRNTSSGQAP-VVLTSNKVGTVTVTASFHNGV 929

RESULT 7  
US-10-793-626-2964  
; Sequence 2964, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P034800S  
; CURRENT APPLICATION NUMBER: US/10793.626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2964  
; LENGTH: 5024  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
; FEATURE:  
; NAME/KEY: MOD\_RES  
; LOCATION: (5024)  
; OTHER INFORMATION: variable amino acid  
US-10-793-626-2964

Query Match 3.3%; Score 164.5; DB 6; Length 5024;  
Best Local Similarity 20.4%; Pred. No. 0.064;  
Matches 208; Conservative 158; Mismatches 407; Indels 249; Gaps 51;

QY 8 LYOKTPTV---SVYDNRGLIIRNIDFHRHTTANGDPDTRITRHOYDIHGLNQSIDPRLYE 64  
DB 7 LVAQPTVQKTSVINE-----DQPSQAYND---SITMGQTIINKTADPVLDTKLVD 56

QY 65 AKQTNVTKPNFL-WQYDLTGNPLCTESIDAGRTVT--LNDIEGRPLLTATGVIQOT--- 119  
DB 57 NAINISITKENALHGEQKLT--TAKTEAINALNTLADLNTPOKEAIKTAINTAHTRTDVT 114

QY 120 -ROYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGTEAEKDH-----NLAGOCVRHY 172  
DB 115 AEQSKANQINSAMHTLQKINISDNESVTNESVI---NAEPEKQHAFTALNNAKEIVNE- 170

QY 173 DTAGVTRLESLSLTGTVLSSQSOLLIDTQEANWTGDNETVQNMMLADDIYTLSTFDATG 232  
DB 171 -----QOATLDANSINQAKAILTTKNA-LDGEQLRRAKENADQEIINTLNL--- 217

QY 233 ALLTQTDAGNIQRLAYDVAGOLGSMWTLKGTQEQVILKSLTYSAAQOKLREEHGNVDI 292  
DB 218 ----TDAQRNSEK-----GLVNSS-----QTRTEVASQLAKAKELNKVMEQLNN--- 257

QY 293 TEYSVEPETQRLIGIKTRRPSDKVLQDLRYEDVPVGNVISIRNDAEATRFWNNOKVMPE 352  
DB 258 -----LNGKQMINSKFPINEDANQOQAYSNAIA---SAEVLK---NKSQNP 300

QY 353 NYTYVDSLYQLISATGREMANIGQSH---QPPSPALPSDNNT--YTNVTRT----- 399  
DB 301 LDKV--TIEQAINNINSAINNLNGEAKLTAKEDAVASINNLGLTSQKTKENQAVNGS 358

QY 400 YTYDRGGLTKIQHSSSPATQNNYTNITVSNRSRAVLST---LTEDPAQVDALPDAGCH 456  
DB 359 QTRDQVANV--LRDSKALDQSMQTLRLDLVNNQN--VIHSTSNYFNEDSTQKNTYDNIADN 414

QY 457 QNTLISQON---LWNTEGE-LQQVTLVKRD-KGANDDREWVRYSGDGRMLKINEQAS 511  
DB 415 GSTVITGQHSBELNKSSTIDQITISQINTAKNDLHGAE-----KLQDRKGT 458





;; CURRENT APPLICATION NUMBER: US/11/013,759  
;; CURRENT FILING DATE: 2004-12-16  
;; PRIOR APPLICATION NUMBER: US/09/361,619  
;; PRIOR FILING DATE: 1999-07-27  
;; NUMBER OF SEQ ID NOS: 32  
;; SOFTWARE: Patencin Ver. 2.1  
;; SEQ ID NO 11  
;; LENGTH: 2314  
;; TYPE: PRT  
;; ORGANISM: Moraxella catarrhalis  
US-11-013-759-11

Query Match 3.1%; Score 156; DB 7; Length 2314;  
Best Local Similarity 19.1%; Pred. No. 0.072;  
Matches 222; Conservative 157; Mismatches 398; Indels 384; Gaps 59;

QY 3 NIDPKLYQKTPVSVDNRG---LIRNIDFHRTTANG-----DPDTRITRHQYDI 50  
DB 814 NVDEK-----TIEUTGONGTKNGIKGVKTTITTTNANGKATNFSTTNDALV--NAXDI 865  
QY 51 HGHNLQSDIPRIYEAQTNNIKPNFLWQYD--LTGNPLCTESIDA---GRTVTLLNDIEGR 106  
DB 866 AENLN--TLAKEHTTKGTADTALQTFKVKDGATDDEITVKGDTQNGKTWNTLKLGE 924  
QY 107 PLLTV-----TATGVIQT---ROYETSSILPGRLLSVABQTPPEK-----142  
DB 925 NGLTVATNKDGTVPFGINTQSLKAGDSTLLKGLSINKPNASNEQIQVGADGVKFAKVD 984  
QY 143 -----TSRITERLIW---AGNTEAEKDH-----NLGQCVR 170  
DB 985 KNSSTGIDGTSRIKQDQIGFTGANGSLDTTKPHLTKDKLVKGEVEIINTGINAGGKIT 1044  
QY 171 HYDTAGVTRLESILT--GTVLSQSSQLLIDTQEAANWTGNETWOMLADD-----IYT 223  
DB 1045 NIQSDITQNSDNDATGVRVLDKTELESKINSAAKTAQN--SLHEFSVADQGNHPTVSN 1103  
QY 224 TLSTFPA--TGALLT-----QTDAKNIQRLAYD-----VAGQLNGSWLTUKG 264  
DB 1104 PYSSYDTSKTSVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTVTGNNGKGIVIDS 1163  
QY 265 QTEQVILKSL-----TYSAGQKLREHGNVDVITEYSYEPETQRLIGIKTRPSDKTV 317  
DB 1164 KDGQNTITGLSNTLANVNDGAGHALSQGLANDT-----DKTRAAS-----1204  
QY 318 LQDLRYEYDVPGNVI-----STRNDAEATRFWHNQKMPENTVYDLSYLQSLSATGREMAN 373  
DB 1205 -----IGDVLNAGFNLCNGEAVDF-----VSTYDTV-----1231  
QY 374 IGQSQHQFPSPALPSDNNY--TNTRYTYDVRGNLTKIQHSSPATQNNVTNITVSNR 431  
DB 1232 -----DFIDGNATTAKVYDDTSKTSKVYDVNDVNDKNTIEVTSKGLGVKTTTLTKTSA 1285  
QY 432 SNRAVLSTLTEDPAQVDALFAGGHONTLISQNLNWNTRGELQVTLVKRDKGAND---488  
DB 1286 NGNATKFSAADGDALVKA--SDIATHLNTL-----AGDIQTA-----KGASQASS 1328  
QY 489 -----DREWRYSGDG-----RRMLK-----504  
DB 1329 SASVYDADGNKVIYDSTDKKYQVNDKQVDKNKEVAKDLVAQAQTQDGLTAQNNVKS 1388  
QY 505 INEQASNNNAQTRV-----TYLPNLELRL-----TONSTATTEDLQVIT-----VGEAGR 550  
DB 1389 INKEQVNDANKQGINEDNAFKGLENAAKDTTKNAATVVDGLNNAVAQTPLTFAGDTGT 1448  
QY 551 AQVRVLHWSGKPEDIDNNOLRYSVDNLIGSSOLEDSGQIISSEYYPYGGTALWAAR 610  
DB 1449 TAKKLGELTITKGGQTDNKLK-----DNNIG-----VWAGTD-----GFTVVKLAK 1489  
QY 611 NOTEASVYKTIYRSGKERATGLYYGYRYQYQWIGRWLSSDPAGTFIDGLNL--YRMVRNPP 669  
DB 1490 DLT--NLNSVNAAGGTIRBEKGISFVDAN-----QOAKANTPVL SANGLDLGGKISNIG 1541  
QY 670 VTLLDPDGL-----MPTIAERIAALKKNK-----VTD SAPSAPANATNVAINRPVA 716

## RESULT 10

US-10-467-657-1088  
; Sequence 1088, Application US/10467657  
; Publication No. US20050260581A1

;; GENERAL INFORMATION:  
;; APPLICANT: CHIRON Spa  
;; APPLICANT: FONTANA Maria Rita  
;; APPLICANT: PIZZA Mariagrazia  
;; APPLICANT: MASIGNANI Vega  
;; APPLICANT: MONACI Elisabetta  
;; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
;; FILE REFERENCE:  
;; CURRENT APPLICATION NUMBER: US/10/467,657  
;; CURRENT FILING DATE: 2003-08-11  
;; PRIOR APPLICATION NUMBER: GB-0103424.8  
;; PRIOR FILING DATE: 2001-02-12  
;; NUMBER OF SEQ ID NOS: 9218  
;; SOFTWARE: SeqWin99, version 1.04  
;; SEQ ID NO 1088  
;; LENGTH: 1468  
;; TYPE: PRT  
;; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-1088

Query Match 3.1%; Score 154.5; DB 6; Length 1468;  
Best Local Similarity 20.6%; Pred. No. 0.046;  
Matches 213; Conservative 130; Mismatches 378; Indels 315; Gaps 55;

QY 15 VSVYDNRGLIIRNIDFHRRTTANGPDTRITRHOYDIHGLHMQSIDPRLYEAKQNTNITKP 74  
DB 111 VSVAHNGS--YNNVDFGAGSNPD-----QHRFSY-----QIVKBNYKAGTNHPIYGG 157  
QY 75 NF-----LWQYDLTGNNPL--CTESIDAGRTVTLLNDIEGRPLLTATGTATQYETSSLP 129  
DB 158 DYHMPRLHFTVDAEPVEMTSYMDGWKYADLNKYPDR-----VRIGAGRYWRSD---207  
QY 130 RLLSVAEQTPPEKTS--RITERLIW--AGNTEAEKDH-----NLAGOCVRHYDVTAGVTRL 180  
DB 208 -----EDEPNRESSVHIASAYSWLVGNGTFAQNGSGGTVNLGSEKIKHSFYGLPTG 261  
QY 181 ELSLITGTVLSQSSQLLIDTQEAANW-----TGDN-----ETVMQNNLADDIYT---223  
DB 262 GSFDSG-----SPWFIYDAQKQKWLNGVLTGNPYIGKSNGLFOLVRKDFWDFDI FADG 316  
QY 224 TLSTF-----DATGALLTQTDKAGNTQRLAYDVAGQLNGSWLTLKGTQEQVI 270  
DB 317 THSVFYEPHQNGKYFFNDNNNGA--GKIDAKHKHVSFPY-----RLKTRTVQLF 363  
QY 271 IKSLTYSAGQKLREHGNVDVITEYSYEPETQRLIGIKTRRPSDKVLQDLRYEYDVPGN 330



```

US-11-113-424-54
; Sequence 54, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-54

Query Match      3.1%; Score 153; DB 7; Length 1045;
Best Local Similarity 19.2%; Pred. No. 0.035;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

Qy 15 VSVYDNR---GLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNN 70
Db 106 VNVFGRKLVNGRNLLSVDFRTTK-----TEKIYDDH-----138

Qy 71 TIKPNFLWQ--YDLTGNPLCTESIDAGRTVTNLNDIEGRPLLTATVATGVIQTOYETSSLP 128
Db 139 ---RKFLRLIAYDTSGHP--TLWLPSSKLMVN-----VTYSSTGQIASIQRTGTS--184

Qy 129 GRLLSVARQTEPEKTSRITERLIWAGNTEAEKDHNLACQCVRHVDYTAGVTRLESLSLTGT 188
Db 185 -----EKVDYDGGQGRIVSRVFDGKT-----WSYT 209

Qy 189 VLSQSSQLLIDTOEANTWGDNETVQNMLADDIYTLTSTFDATGALLTQTDKAGNIQRLA 248
Db 210 VLEKSMVLLHSQR-----QYIFEDMDRLS-----AITWPSVARHTMQ---249

Qy 249 YDVAGQLNGSMLTLKGTEQVILKSLTYSAGQKLRBEHGNVDITEYSYEP-----E 300
Db 250 -----TIRSIGYRNIYNPPESNAS-IITDYNEEGLLQTAFLG 287

Qy 301 TORLIGIKTRRPSD--TKVLQD---LRVEYDPVGNVISIRN-----D 337
Db 288 TSSRVLFKYRQTRLSILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLID 347

Qy 338 AEATRFHNOVKMPENTYDLSYLQISATG-----REMANIGQOSHQP-----382
Db 348 RQIFPSEGGWNAFDSYDNSFRVTSWQGVINETPLIDYQFDDISGRVQEGKEGV 407

Qy 383 -----SPALPSDNNITYN-----YTRTTYTDRCGNLTKEIHSSPA 417
Db 408 IYYDINOIISTAVMTYTKHFDAGHRIKEIQEYIFRSLMWITIQIDNMGVTRKEIKGP 467

Qy 418 TON-----NY-----TTNITVSNRNRVLSLTLED-----443

```

```

Db 468 PANTTKYAYEYVDVGQQTQVYVNEKIMWRNYNDLNGNLHLLNPSNSARLTPLRYDLRDRI 527
Qy 444 --PAQVDALFDAGG--HQNTLISQNLNMTWRELQOQVTLVKRDKGANDREW---YRY 495
Db 528 TRLGDVQVRLDEDEGFLRQGTETI-----FEYSKGLLTRV-----YSKSGS---WTVIYRY 575
Qy 496 SGDCGRMLKINEQOASNNATQ-----RVTYLPNLRLTQNSTATTEDLQVITV 545
Db 576 DGLGRV-----SSKTSLGQHLQFPYADLTYPTRITHVYN-----HSSSETLSYYDLQ---624
Qy 546 GEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSBQIISBEEYYPGGTA 605
Db 625 -----GHLFAMEISSG-----DEFYIASDN-TGTPLAVFSSNGLMLKQIQTAYGEIY 671
Qy 606 LWAARNQTEASYKTI--RYSCKERDAGLYYYGYRYQYQWIGRW-----LSSDPA 653
Db 672 F-----DSNIDFQLVIGFHGGLYDPLTKLIHFGERDYLILAGRWTTTPDIEIKRIGKDP 726
Qy 654 GTIDGLNLYRMVRNPNVT 671
Db 727 ----PFNLYMFRNNPNAS 740

RESULT 13
US-10-821-234-1097
; Sequence 1097, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; FILE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1097
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1097

Query Match      3.1%; Score 153; DB 6; Length 1094;
Best Local Similarity 19.2%; Pred. No. 0.037;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

Qy 15 VSVYDNR---GLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNN 70
Db 155 VNVFGRKLVNGRNLLSVDFRTTK-----TEKIYDDH-----187

Qy 71 TIKPNFLWQ--YDLTGNPLCTESIDAGRTVTNLNDIEGRPLLTATVATGVIQTOYETSSLP 128
Db 188 ---RKFLRLIAYDTSGHP--TLWLPSSKLMVN-----VTYSSTGQIASIQRTGTS--233

Qy 129 GRLLSVARQTEPEKTSRITERLIWAGNTEAEKDHNLACQCVRHVDYTAGVTRLESLSLTGT 188
Db 234 -----EKVDYDGGQGRIVSRVFDGKT-----WSYT 258

Qy 189 VLSQSSQLLIDTOEANTWGDNETVQNMLADDIYTLTSTFDATGALLTQTDKAGNIQRLA 248
Db 259 VLEKSMVLLHSQR-----QYIFEDMDRLS-----AITWPSVARHTMQ---298

Qy 249 YDVAGQLNGSMLTLKGTEQVILKSLTYSAGQKLRBEHGNVDITEYSYEP-----E 300
Db 299 -----TIRSIGYRNIYNPPESNAS-IITDYNEEGLLQTAFLG 336

Qy 301 TORLIGIKTRRPSD--TKVLQD---LRVEYDPVGNVISIRN-----D 337
Db 337 TSSRVLFKYRQTRLSILYDSTRVSFTYDETAGVLKTVNLQSDGFICTIRYRQIGPLID 396

```



; LENGTH: 2715  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-096-051-2

```
Query Match          3.1%; Score 153; DB 7; Length 2715;
Best Local Similarity 19.2%; Pred. No. 0.14;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

QY 15 VSVYDNR---GLIIRNIDFHRTTANGDPDTRTRHQYDIHGHNLQSIDPRLYEAKQTNN 70
Db 1776 VNVFGKLRVNGRNLLSVDFTTK-----TEKIYDDH-----1808

QY 71 TIKPNFLWQ--YDLFGNPLCTESIDAGRTVTIANDIEGRPLLVTAATGVIQROYETSSLP 128
Db 1809 ---RKFLRLIAYDTSGHP--TLWLPSSKLMVAVN-----VTYSSTGQIASIQRTGTS--1854

QY 129 GRLLSVAQTPPEKTSRTERLIWAGNTEAEKDHNLACQCVRHDTAGVTRLESLSLTGT 188
Db 1855 -----EKVDYDGGQGRIVSRVFDGKT-----WSYT 1879

QY 189 VLSQSSQLLIDTQEANWTGDNETVQNMMLADDIYTLSTFDATGALLTQTDAGNIQRLA 248
Db 1890 YLEKSMVLLHSQR-----QYIFEDMDRLLS-----AITMPSVARHTMQ---1919

QY 249 YDVAGQLNGSWLTKGQTEQVILIKSLTYSAGQKREEHGNDVITEYSYEP-----E 300
Db 1920 -----TIRSIGYRYNIYNPESNAS-IITDYNEEGLLLQTAFLG 1957

QY 301 TORLIGIKTRPSD--TKVLQD---LRYEYDPVGNVISIRN-----D 337
Db 1958 TSRRVLFKYRQTRLSILYDSTRVSFTYDAGVLKTVNLQSDGFICTIRYRQIGPLID 2017

QY 338 AEATFENQKVPENTTYDSLYQLISATG-----REMANIGQOSHOPP---382
Db 2018 KQIFRPSDGMVNAFSDYSNFSRVTSMQGVINETPLPIDLYQFDDISGRVQEGKFGV 2077

QY 383 -----SPALPSDNNTYTN-----YTRTYTYDRGNNLTQIHHSPA 417
Db 2078 IYVDINQIISTAVMTYTKHFDAGHRIKEIQYEIIFRSLMWITTIQYDNNMGRTKREIKIGP 2137

QY 418 TQN-----NY--TTNITVSNRRAVLSTLTED-----443
Db 2138 FANTTKYAYEYVDGQLOTVYLNKIMWRNYNDLNGNLHLLNPSNSARLTPLRYDLDRDI 2197

QY 444 --PAQVDALFDAGG---HQNTLISQNLNWNTRGELQOVTLVKRDKGANDDREW---YRY 495
Db 2198 TRLGDVQVRLDEDFLQRGTEI-----FEYSKGLLTV-----YSKSGS---WTVIYRY 2245

QY 496 SGDGRMLKINEQQAASNAQTQ-----RVTYLPNLELRLTQNSTATTEDLQVITV 545
Db 2246 DGLGRV-----SSKTSLGHLQFFYADLTYPTRITHVYN---HSSSEITSLYDQLQ---2294

QY 546 GEAGRAQVRVLHWESGKPEDIDNNQRYSDYNLIGSSQLELDSQGLISBEEYYPYGTA 605
Db 2295 -----GHLFAMEISSG-----DEFYIASDN-TGTPLAVFSSNGLMLKIQYAYGEIY 2341

QY 606 LWAARNOFEASYKTI--RYSKGERDATGLYYGYGYQYQPWIGRW-----LSSDPA 653
Db 2342 F-----DSNIDFQVLGHGGLYDPLTKLIHFGERDYLILAGRWTTTPDIEIWKRIKQDPA 2396

QY 654 GTIDGLNLYRMVRNPNVT 671
Db 2397 -----PFNLYMFRNNPNAS 2410
```

Search completed: February 16, 2006, 21:55:22  
Job time : 12.1786 secs

**THIS PAGE BLANK (USPTO)**

---

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 95.1891 Seconds  
(without alignments)  
4431.220 Million cell updates/sec

Title: US-10-754-115-47

Perfect score: 5005

Sequence: 1 MKNDPKLYQKTPVSYVDN.....DAEISFLTTLPLKNVXPKHR 960

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5005	100.0	960	ADP18624	Adp18624 Photorhab
2	5005	100.0	960	ADR21527	Adr21527 Photorhab
3	4545	90.8	954	AAY33728	Aay33728 Photorhab
4	3300.5	65.9	971	Abm70223	Abm70223 Photorhab
5	2879	57.5	938	ADP18628	Adp18628 Photorhab
6	2879	57.5	938	ADR21579	Adr21579 Photorhab
7	2850.5	57.0	949	ADP18626	Adp18626 Photorhab
8	2850.5	57.0	949	ADR21544	Adr21544 Photorhab
9	2833.5	56.6	939	ABM70226	Abm70226 Photorhab
10	2821.5	56.4	960	ABM70214	Abm70214 Photorhab
11	2768.5	55.3	966	ABM69154	Abm69154 Photorhab
12	2744	54.8	1044	ABM69055	Abm69055 Photorhab
13	2740	54.7	1043	AAW18306	Aaw18306 Photorhab
14	2740	54.7	1043	AAW56569	Aaw56569 Toxin Tcc
15	2740	54.7	1043	ABM70231	Abm70231 Photorhab
16	2539	50.7	916	ABG32853	Abg32853 P. lumine
17	2537.5	50.7	915	ABG32853	Abg32853 P. lumine
18	2535.5	50.7	915	ADR21538	Adr21538 Photorhab
19	2383	47.6	937	ABM67982	Abm67982 Photorhab
20	2305	46.1	971	AAY95687	Aay95687 Cosmid cH
21	2304.5	46.0	962	ADR21531	Adr21531 Xenorhabd
22	2304.5	46.0	962	ABE47814	Aeb47814 Native Xp
23	2226	44.5	1016	ADR20367	Adr20367 Recombina
24	2226	44.5	1016	ADR21496	Adr21496 Xenorhabd

25	2184	43.6	973	4	AAY97696	Aay97696 SepC prot
26	2184	43.6	973	8	ADR21541	Adr21541 Serratia
27	1581.5	31.6	930	8	ADN61349	Adn61349 Paenibaci
28	1581.5	31.6	930	8	ADR21522	Adr21522 Paenibaci
29	1581.5	31.6	930	8	ADN61355	Adn61355 Paenibaci
30	1581.5	31.6	953	8	ADR21523	Adr21523 Paenibaci
31	1268.5	25.3	940	7	ADG73071	Adg73071 Pseudomon
32	1268.5	25.3	940	7	ADL12126	Adl12126 Pseudomon
33	651	13.0	286	8	ADN61377	Adn61377 Paenibaci
34	332.5	6.6	932	6	ABU16693	Abu16693 Protein e
35	329.5	6.6	1572	6	ABU41491	Abu41491 Protein e
36	329.5	6.6	1627	6	ADA35317	Ada35317 Acinetoba
37	322.5	6.4	998	7	ABO79678	AbO79678 Pseudomon
38	321	6.4	1530	6	ABU40251	Abu40251 Protein e
39	318	6.4	1385	6	ABU40318	Abu40318 Protein e
40	317	6.3	1565	6	ABU16634	Abu16634 Protein e
41	315	6.3	1317	8	ADS22210	Ads22210 Bacterial
42	308	6.2	2334	5	ABG31849	Abg31849 Human kin
43	307.5	6.1	843	6	ABU48290	Abu48290 Protein e
44	303.5	6.1	1439	7	ADF07275	Adf07275 Bacterial
45	302.5	6.0	1253	6	ABU40990	Abu40990 Protein e

## ALIGNMENTS

### RESULT 1

ADP18624

ID ADP18624 standard; protein; 960 AA.

XX AC ADP18624;

XX DT 12-AUG-2004 (first entry)

XX DE Photorhabdus luminescens W-14 tccC3 protein SEQ ID NO:12.

XX KW tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

XX KW orally active insect toxin; insect toxin; tccc3.

XX OS Photorhabdus luminescens.

XX PN WO2004044217-A2.

XX PD 27-MAY-2004.

XX PF 12-NOV-2003; 2003WO-IB005553.

XX PR 12-NOV-2002; 2002US-0425672P.

XX (UYBA-) UNIV BATH.

XX Ffrench-Constant RH, Waterfield NR;

XX WPI; 2004-411735/38.

XX N-PSDB; ADP18623.

XX New isolated Photorhabdus luminescens nucleic acids, useful for expressing orally active insect toxin or for generating transgenic plants with enhanced resistance to insects.

XX Claim 1; SEQ ID NO 12; 118pp; English.

XX The present invention describes DNA sequences from the tcd genomic region of Photorhabdus luminescens W-14. Also described: (1) a transgenic monocot or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of Photorhabdus luminescens W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to

CC insects. The present sequence represents Photorhabdus luminescens W-14  
CC tccc3, which is used in the exemplification of the present invention.  
XX  
SQ Sequence 960 AA;

Query Match 100.0%; Score 5005; DB 8; Length 960;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60  
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60

Qy 61 RLYEAKQNTNNTIKNFWLQYDLTGMPCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
Db 61 RLYEAKQNTNNTIKNFWLQYDLTGMPCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180  
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180

Qy 181 ESLSUTGTVLSCSSQLLIDTQGANWGTGNETWQNMADDIYTTLTSTPDATGALLTQDPA 240  
Db 181 ESLSUTGTVLSCSSQLLIDTQGANWGTGNETWQNMADDIYTTLTSTPDATGALLTQDPA 240

Qy 241 KGNITQRLAYDVAGQLNGSWLTLKGTEOVIIKSLTYSAGOKLREHGNNDVITEYSVEPE 300  
Db 241 KGNITQRLAYDVAGQLNGSWLTLKGTEOVIIKSLTYSAGOKLREHGNNDVITEYSVEPE 300

Qy 301 TQRLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTYTYDSL 360  
Db 301 TQRLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTYTYDSL 360

Qy 361 YQLISATGREMANIQQSHQSPALPSDNNYTYTFTYTYDRGNNLTKIQHSPATON 420  
Db 361 YQLISATGREMANIQQSHQSPALPSDNNYTYTFTYTYDRGNNLTKIQHSPATON 420

Qy 421 NYTTNITVSNRNLAVLTSTLTPDPAQVDALPDAGHONTLSGQNLNNTREGELQOVTLV 480  
Db 421 NYTTNITVSNRNLAVLTSTLTPDPAQVDALPDAGHONTLSGQNLNNTREGELQOVTLV 480

Qy 481 KRDKGANDREWIRYSGDGRMLKINEQQAASNAQTRVYLPNLERLATQNSTATTEDL 540  
Db 481 KRDKGANDREWIRYSGDGRMLKINEQQAASNAQTRVYLPNLERLATQNSTATTEDL 540

Qy 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQRLYSYDNLIGSSQLELDSGQIISSEYYP 600  
Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQRLYSYDNLIGSSQLELDSGQIISSEYYP 600

Qy 601 YGGTALWAARQNTASQYKTIYSGKERDATGLYTYGYRYOPWIGRWLSSDPAGTIDGLN 660  
Db 601 YGGTALWAARQNTASQYKTIYSGKERDATGLYTYGYRYOPWIGRWLSSDPAGTIDGLN 660

Qy 661 LYRVRNPNVTLDPDGLMPTIAERIAALKKNKVYTDSPANATVAINIRPPVAPKPS 720  
Db 661 LYRVRNPNVTLDPDGLMPTIAERIAALKKNKVYTDSPANATVAINIRPPVAPKPS 720

Qy 721 LPKASTSSQPTTHPIGAANIPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSTTST 780  
Db 721 LPKASTSSQPTTHPIGAANIPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSTTST 780

Qy 781 NLQKKSFTLYRADNRSFEEMSKPEPGKATPLDTKWARPASIFIGQKDTSNLPKETV 840  
Db 781 NLQKKSFTLYRADNRSFEEMSKPEPGKATPLDTKWARPASIFIGQKDTSNLPKETV 840

Qy 841 KNIISTWGAQPKLKDLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
Db 841 KNIISTWGAQPKLKDLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900

Qy 901 GQKLNPLPEGRTKNVPSLLDTPQIETSSIIALNHGVPNDABISFLTITPLKNVKKPHR 960  
Db 901 GQKLNPLPEGRTKNVPSLLDTPQIETSSIIALNHGVPNDABISFLTITPLKNVKKPHR 960

RESULT 2  
ADR21527  
ID ADR21527 standard; protein; 960 AA.  
XX  
AC ADR21527;  
XX  
DT 04-NOV-2004 (first entry)  
XX  
DE Photorhabdus Tccc3 toxin SEQ ID NO:47.  
XX  
KW toxin; insect; insecticidal; transgenic; pest control.  
XX  
OS Photorhabdus luminescens.  
XX  
FN WO2004067727-A2.  
XX  
PD 12-AUG-2004.  
XX  
PF 07-JAN-2004; 2004WO-US000394.  
XX  
PR 21-JAN-2003; 2003US-0441723P.  
XX  
FA (DOWC ) DOW AGROSCIENCES LLC.  
XX  
PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;  
PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;  
XX  
DR WPI; 2004-580999/56.  
DR N-ESDB; ADR21526.  
XX  
PT Controlling or inhibiting an insect, useful for pest control, comprises  
PT contacting the insect with effective amounts of a Protein A, a Protein B,  
XX and a Protein C.  
XX  
PS Claim 1; SEQ ID NO 47; 368pp; English.  
XX  
CC The invention relates to a novel method for controlling or inhibiting an  
CC insect comprising contacting the insect with effective amounts of a  
CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
CC is encoded by a naturally occurring gene or has an amino acid sequence  
CC that differs from the product encoded by a naturally occurring gene only  
CC by truncation or by conservative amino acid changes. Protein A is a 230-  
CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
CC species, has stand alone insecticidal activity, and has an amino acid  
CC sequence at least 40% identical to a sequence selected from XptAlwi,  
CC XptA2w1, TcdA, TcdA2, TcdA4, and TcdBA. Protein B is a 130-180 kDa toxin  
CC complex potentiator having an amino acid sequence at least 40% identical  
CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1w1, XptB1xb,  
CC PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
CC having an amino acid sequence at least 35% identical to a sequence  
CC selected from Tccc1, Tccc2, Tccc3, Tccc4, Tccc5, XptB1w1, XptC1xb, PptC1  
CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
CC C. The method is useful for pest control. The present sequence represents  
CC Photorhabdus luminescens Tccc3 toxin.  
XX  
SQ Sequence 960 AA;

Query Match 100.0%; Score 5005; DB 8; Length 960;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60  
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60

Qy 61 RLYEAKQNTNNTIKNFWLQYDLTGMPCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
Db 61 RLYEAKQNTNNTIKNFWLQYDLTGMPCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180  
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180



Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 180  
 QY 181 ESSLTGTVLSSQSSQLLIDTQEAANTGNQETVWQNLADDIYTTLSFTDAGTALLTQTD 240  
 Db 181 ESSLTGTVLSSQSSQLLIDTQEAANTGNQETVWQNLADDIYTTLSFTDAGTALLTQTD 240  
 QY 241 KGNITQRLAYDVAGQINGSLWTLKGTEQVVIKSLTYSAGQKLRBEHNDVITEYSPE 300  
 Db 241 KGNITQRLAYDVAGQINGSLWTLKGTEQVVIKSLTYSAGQKLRBEHNDVITEYSPE 300  
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWQNKMPENTTYDLS 360  
 Db 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWQNKMPENTTYDLS 360  
 QY 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420  
 Db 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420  
 QY 421 NYTTNITVSNRNRVAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGLOQVTLV 480  
 Db 421 NYTTNITVSNRNRVAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGLOQVTLV 480  
 QY 481 KDKGANDREWRYSGDGRMLKINEQOASNAQTORVYLPNLELRLTONSTATTEDL 540  
 Db 481 KDKGANDREWRYSGDGRMLKINEQOASNAQTORVYLPNLELRLTONSTATTEDL 540  
 QY 541 QVITVGEAGRAQVRVHWSGKPEDIDNNQLRYSYDNLIGSSQLDSEGQIISBEEYYP 600  
 Db 541 QVITVGEAGRAQVRVHWSGKPEDIDNNQLRYSYDNLIGSSQLDSEGQIISBEEYYP 600  
 QY 601 YGGTALWAARQTEASYKTIYSGKERDATGLYYGYRYQPWIGRWLSSDPAGTIDGLN 660  
 Db 601 YGGTALWAARQTEASYKTIYSGKERDATGLYYGYRYQPWIGRWLSSDPAGTIDGLN 660  
 QY 661 LYRWVRNPNVTLDPDGLMPTTAERIALKKNKVTDSPSPANATVAINIRPPVAPKPS 720  
 Db 661 LYRWVRNPNVTLDPDGLMPTTAERIALKKNKVTDSPSPANATVAINIRPPVAPKPS 720  
 QY 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTTST 780  
 Db 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTTST 780  
 QY 781 NLQKSFLLYRADNRSFEMQSKFPEGFKAWTPLDTKMAROPASIFIGQKOTSNLPKETV 840  
 Db 781 NLQKSFLLYRADNRSFEMQSKFPEGFKAWTPLDTKMAROPASIFIGQKOTSNLPKETV 840  
 QY 841 KNIISTWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 Db 841 KNIISTWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900  
 QY 901 GQKLNPLPEGRTKNMPVSLLDTPQIETSSIIALNHPVNDABISFLTTIPLKNVPHKR 960  
 Db 901 GQKLNPLPEGRTKNMPVSLLDTPQIETSSIIALNHPVNDABISFLTTIPLKNVPHKR 960

RESULT 3

AAV33728  
 ID AAV33728 standard; protein; 954 AA.  
 XX  
 AC AAV33728;  
 XX  
 DT 09-NOV-1999 (first entry)  
 XX  
 DE Photorhabdus luminescens 954 amino acid insecticidal toxin.  
 XX  
 KW Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.  
 XX  
 OS Photorhabdus luminescens.  
 XX  
 PN WO9942589-A2.  
 XX  
 PD 26-AUG-1999.  
 XX

PF 18-FEB-1999; 99WO-EP001015.  
 XX  
 PR 20-FEB-1998; 98US-00027080.  
 PR 20-JAN-1999; 99US-0116439P.  
 XX  
 PA (NOVS ) NOVARTIS AG  
 PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
 XX  
 PI Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM;  
 PI Chen JS;  
 XX  
 DR WPI; 1999-527479/44.  
 DR N-PSDB; AA06831.  
 XX  
 PT New nucleic acid from Photorhabdus luminescens encoding insecticidal  
 PT toxins, used for making resistant transgenic plants.  
 XX  
 PS Claim 26; Page 130-133; 148pp; English.  
 XX  
 CC This sequence represents a 954 amino acid insecticidal toxin from  
 CC Photorhabdus luminescens. It is one of three insecticidal toxins  
 CC (AA03728-Y33730) encoded by open reading frames (orfs) in a 38kb  
 CC fragment of P. luminescens DNA (AA06831). This sequence is encoded by  
 CC orf5. P. luminescens is a member of the Enterobacteriaceae family and is  
 CC a symbiotic bacterium of nematodes of the genus Heterorhabditis. The  
 CC nematodes colonise insect larvae, kill them, and their offspring feed on  
 CC the dead larvae. However, the insecticidal agents are produced by P.  
 CC luminescens rather than the nematodes. The toxins have activity against  
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni), European  
 CC Corn Borer (Ostrinia nubilalis) and Fall Armyworm (Spodoptera frugiperda)  
 CC and also against Coleopteran insects (e.g., Colorado Potato Beetle,  
 CC Leptinotarsa decimlineata). In addition the toxins are active against  
 CC strains resistant to known insecticides. The DNA sequence can be used to  
 CC generate transgenic plants of various species that are resistant to  
 CC economically important insect pests and also for recombinant production  
 CC of the toxins for use as insecticides  
 XX  
 SQ Sequence 954 AA;

Query Match 90.8%; Score 4545; DB 2; Length 954;  
 Best Local Similarity 90.4%; Pred. No. 3.6e-307;  
 Matches 868; Conservative 40; Mismatches 46; Indels 6; Gaps 2;  
 QY 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHQYDIHGLHNSIDP 60  
 Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHQYDIHGLHNSIDP 60  
 QY 61 RLYEAKQNTNNTIKNPLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120  
 Db 61 RLYEAKQNTNNTIKNPLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 115  
 QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 180  
 Db 116 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 174  
 QY 181 ESSLTGTVLSSQSSQLLIDTQEAANTGNQETVWQNLADDIYTTLSFTDAGTALLTQTD 240  
 Db 175 ESSLTGTVLSSQSSQLLIDTQEAANTGNQETVWQNLADDIYTTLSFTDAGTALLTQTD 234  
 QY 241 KGNITQRLAYDVAGQINGSLWTLKGTEQVVIKSLTYSAGQKLRBEHNDVITEYSPE 300  
 Db 235 KGNITQRLAYDVAGQINGSLWTLKGTEQVVIKSLTYSAGQKLRBEHNDVITEYSPE 294  
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWQNKMPENTTYDLS 360  
 Db 295 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWQNKMPENTTYDLS 354  
 QY 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420  
 Db 355 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 414  
 QY 421 NYTTNITVSNRNRVAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGLOQVTLV 480

Db 415 NYTNTITVSNRNRNAVLTSTLTPDPAQVDALFDAGGHQNTLISGQNLNWNTRGELQHVTLV 474

Qy 481 KRDGANDREWYRSGRRMLKINEQOASNAQTRVTVLPNLELRLTONSTATTEDL 540

Db 475 KRDGANDREWYRSGRRMLKINEQOASNAQTRVTVLPNLELRLTONSTATTEDL 534

Qy 541 QVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEYYP 600

Db 535 QVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEYYP 594

Qy 601 YGGTALWAARNOTEASQKTIYRSGKRDATGLYYGYRYYPQWIGRLSSDPAGTIDGLN 660

Db 595 YGGTALWAARNOTEASQKTIYRSGKRDATGLYYGYRYYPQWIGRLSSDPAGTIDGLN 654

Qy 661 LYRVNRNPNVTLDPDGLMPTIAERIAALKKNKYDTSAPSANATVAINIRPPVAKPS 720

Db 655 LYRVNRNPNVTLDPDGLMPTIAERIAALKKNKYDTSAPSANATVAINIRPPVAKPS 714

Qy 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSTTST 780

Db 715 LPKASTSQPTTPIKASIKPTTSGSSITAPLSPVGNKSTPEISLPSAQSSSTTST 774

Qy 781 NLQKGSFTLYRADNRSEEMQSKPEGPKAWTPLDTKMARQFASIFIGQKDTSNLPKXTV 840

Db 775 NLQKGSFTLYRADNRSEEMQSKPEGPKAWTPLDTKMARQFASIFIGQKDTSNLPKXTV 834

Qy 841 KNIISTWGAKKPLKDLNVIKYTKDSTVWVSTAINTEAGGSSGAPLHKIDMDLYEPAID 900

Db 835 KNIINTWGTGPKLNDLSTVIKYTKDSTVWVSTAINTEAGGSSGAPLHKIDMDLYEPAID 894

Qy 901 GOKNPLPEGTKNWPSLLDTPQIETSSIIALNHGPNVDAEISFLTTIPLKNVKKPHKR 960

Db 895 GOKNPLPRGSKORVPSLLDTPQIETSSIIALNHGPNVDAEISFLTTIPLKNVKKPHKR 954

RESULT 4

ABM70223

ID ABM70223 standard; protein; 971 AA.

XX AC ABM70223;

XX DT 20-NOV-2003 (first entry)

XX DE Photobabidus luminescens protein sequence #3320.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

KW detection; food; gene expression; plant; animal; microorganism; toxin;

KW antibiotic; biopesticide; virulence factor; disease model; plague;

XX whooping cough.

XX OS Photobabidus luminescens.

XX WO200294867-A2.

XX PN 07-FEB-2002; 2002WO-IB003040.

XX PD 28-NOV-2002.

XX PF 07-FEB-2001; 2001FR-00001659.

XX PR (INSP ) INST PASTEUR.

XX PA (CNRS ) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

PI Buchrieser C;

XX WPI; 2003-148459/14.

XX PT Genomic sequence of Photobabidus luminescens and encoded polypeptides,

PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 3320; 1205pp; French.

XX

CC The invention relates to the isolation of genes and their encoded

CC proteins from Photobabidus luminescens. The isolated sequences are

CC sources of probes and primers for detecting the genome of P. luminescens

CC and related species; to study polymorphisms; for gene analysis and for

CC detection/amplification of the genes. Antibodies (Ab) raised against the

CC polypeptides encoded by the genes are used for detection/identification

CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that

CC modulate, regulate, induce or inhibit expression of the genes in plants,

CC animals or microorganisms other than P. luminescens and are able to alter

CC response or sensitivity to toxins and antibiotics produced by P.

CC luminescens. Cells transformed to express the genes are useful for

CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes, proteins, vectors containing the genes and Ab are also useful

CC therapeutically (to treat microbial infection by bacteria or fungi that

CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as

CC biopesticides. Other uses of the genes and the proteins are as virulence

CC factors and for identifying targets of human diseases for which P.

CC luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens proteins

XX

SQ Sequence 971 AA;

Query Match 65.9%; Score 3300.5; DB 6; Length 971;

Best Local Similarity 67.6%; Pred. No. 1.7e-220;

Matches 679; Conservative 81; Mismatches 151; Indels 93; Gaps 17;

Qy 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPDTRITRHOYDIGHLNQSIDP 60

Db 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPDTRITRHOYDIGHLNQSIDP 60

Qy 61 RLYEAKTNTIKENFLWOYDLTGNIPLCTESIDAGRTVTLNDIEGRPLLTVTATGVLTQR 120

Db 61 RLYEAKTNTIKENFLWOYDLTGNIPLCTESIDAGRTVTLNDIEGRPLLTVTATGVLTQR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGACVRRHYDTAGVTRL 180

Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGACVRRHYDTAGVTRL 180

Qy 181 ESSLTGTVLSSQSLLDITQENWGTGNETWQWMLADDIYTLTSTFDATGALLTQTD 240

Db 181 ESSLTGTVLSSQSLLDITQENWGTGNETWQWMLADDIYTLTSTFDATGALLTQTD 240

Qy 241 KGNIRLAYDVAGOLNGSLTLKGTQVLIKSUTYSAAGOKLREHNDVITEYSYEPE 300

Db 241 KGNIRLAYDVAGOLNGSLTLKGTQVLIKSUTYSAAGOKLREHNDVITEYSYEPE 300

Qy 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDABATRFHNNQKVPENTY 355

Db 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDABATRFHNNQKVPENTY 360

Qy 356 TYDSLYQLISATGREMANIGQSHQFPSPALPSDNNYTYTNTYTRTYTDRGNLTKIQHSS 415

Db 361 IYDSLYQLISATGREMANIGQSHQFPSPALPSDNNYTYTNTYTRTYTDRGNLTKIQHSS 420

Qy 416 PATQNNYTNITVSNRNRNAVLTSTLTPDPAQVDALFDAGGHQNTLISGQNLNWNTRGELQ 475

Db 421 PATQNNYTNITVSNRNRNAVLTSTLTPDPAQVDALFDAGGHQNTLISGQNLNWNTRGELQ 480

Qy 476 QVTLVKRDKGANDREWYRSGRRMLKINEQOASNAQTRVTVLPNLELRLTONSTATTEDL 535

Db 481 QVTLVKRDKGANDREWYRSGRRMLKINEQOASNAQTRVTVLPNLELRLTONSTATTEDL 540

Qy 536 TTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEE 595

Db 541 TTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEE 600

Qy 596 BEYYPYGGTALWAARNOTEASQKTIYRSGKRDATGLYYGYRYYPQWIGRLSSDPAGT 655

Db 601 BEYYPYGGTALWAARNOTEASQKTIYRSGKRDATGLYYGYRYYPQWIGRLSSDPAGT 660

Qy 656 IDGLNLYRVNRNPNVTLDPDGLMPTIAE-----RIAAALKKNKYDTS 697



DB 895 PE---VVSMKH 902

RESULT 6

ADR21579

ID ADR21579 standard; protein; 938 AA.

XX ADR21579;

AC 04-NOV-2004 (first entry)

DT Photorhabdus Tcc5 toxin.

DE toxin; insect; insecticidal; transgenic; pest control.

KW Photorhabdus luminescens; W14.

OS Photorhabdus luminescens; W14.

XX WO2004067727-A2.

XX 12-AUG-2004.

XX 07-JAN-2004; 2004WO-US000394.

XX 21-JAN-2003; 2003US-0441723P.

XX (DOWC) DOW AGROSCIENCES LLC.

XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;

PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;

PI WPI; 2004-580999/56.

XX N-PSDB; ADR21537.

DR Controlling or inhibiting an insect, useful for pest control, comprises

PT contacting the insect with effective amounts of a Protein A, a Protein B,

PT and a Protein C.

XX Claim 1; SEQ ID NO 57; 368pp; English.

XX The invention relates to a novel method for controlling or inhibiting an

CC insect comprising contacting the insect with effective amounts of a

CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C

CC is encoded by a naturally occurring gene or has an amino acid sequence

CC that differs from the product encoded by a naturally occurring gene only

CC by truncation or by conservative amino acid changes. Protein A is a 230-

CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic

CC species, has stand alone insecticidal activity, and has an amino acid

CC sequence at least 40% identical to a sequence selected from XptAlwi,

CC XptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin

CC complex potentiator having an amino acid sequence at least 40% identical

CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptC1xb,

CC PptB1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator

CC having an amino acid sequence at least 35% identical to a sequence

CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptC1wi, XptC1xb, PptC1

CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic

CC plant or plant cell that produces a Protein A, a Protein B, and a Protein

CC C. The method is useful for pest control. The present sequence represents

CC Photorhabdus luminescens Tcc5 toxin.

XX Sequence 938 AA;

Query Match 57.5%; Score 2879; DB 8; Length 938;

Best Local Similarity 60.9%; Pred. No. 3.7e-191;

Matches 591; Conservative 90; Mismatches 186; Indels 104; Gaps 16;

QY 1 MKNIDPKLYQKFTVSVYDNRGLIIRNIDFHRITANGDPDTRITRHOYDIHGLNQSIDP 60

DB 1 MENIDPKLYHTPTVSVHDNRGLAIRNISFHRITAEANTDIRITRHOYNAGLYNQSIDP 60

QY 61 RLYAKQNTNKNFLKQYDITGNPLCTESIDAGRTVTLNDIEGRPLLTWTATGVIQTR 120

DB 61 RLYDAKQNTNNAQVNFIRWHLNLTGNILRTESVDAGRTITLNDIEGRPLLTNAAGVRQNH 120

QY 121 QYETSLPGRLLSVAEO--TPEKTSRITERLIWAGNTEAEKDHNLACQCVRHVYDTAGVTR 179

DB 121 RYEDNTLPGRLLALISEQQAEEKI--TERLIWAGNTPQEKDHNLACQCVRHVYDTAGLTQ 177

QY 180 LESLSLTGTVLSQSSQLLIDTQEBANWTGDNETWQNMMLADDIYTTLTSTFDATGALLTQTD 239

DB 178 LNSLALTGAVLSSQSLTDNQADMTGEDQSLWQKLSDDVYITQSTNTDATGALLTQTD 237

QY 240 AKGNIQRLAYDVAGOLNGSWLTLKGOTEQVLIISLTYSAAGOKLREEHGNDVITEYSVEP 299

DB 238 AKGNIQRLAYDVAGOLNGSWLTLKGQAEQVLIISLTYSAAGOKLREEHGNDVITEYSVEP 297

QY 300 ETORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATREHNNKQVMPENTYTYDS 359

DB 298 ETQELIGITTRRPSDAKVLQDLRYQYDPVGNVISIRNDAEATREHNNKQVMPENTYTYDS 357

QY 360 LYQLISATGREMANIGQSHQFPSPALPSDNNTVNTYTRTYTYDRGNLTIKIQUSSPATQ 419

DB 358 LYQLISATGREMANIGQSNQLPSPALPSDNNTVNTYTRTYTYDRGNLTIKIQUSSPAAQ 417

QY 420 NNYTNTITVNSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISGNLNNWNTRGELQOVTL 479

DB 418 NNYTTDTITVNSNRNRAVLSTLTADPTQVDALFDAGGHQNTLISGNLNNWNTRGELQOVTL 474

QY 480 VKRDKGANDD--REWYRYSGDGRMLKINEQQAASNNAAQTQRTVYLPNLELTLQNTATT 537

DB 475 -----ANNSAGNENWRYDSNGIRQLKNEQQTQNIPOQQRVTVLPGLEIRTTQNNATT 528

QY 538 EDLOVITVGEAGRAQVRLVHWSGKPEDINNNQIRSYDNLIGSSQLELSEGOIISBEE 597

DB 529 EELHVTILGKAGRAQVRLVHWSGKPEDINNNQIRSYDNLIGSSQLELSEGOIISBEE 588

QY 598 YYPYGGTALWAARNQTEASVYKTRYSCKERDATGLYYGYRYQYQWIGRWLSSDPAGTID 657

DB 589 YYPYGGTALWAARNQTEASVYKTRYSCKERDVTLGYTYGYRYQYQWIGRWLSDPAGTID 648

QY 658 GLNLYRMVRNPNVTLDPDGLMPTIABERIAALKKNKVTDSAPSAPANATVAI-----N 710

DB 649 GLNLYRMVRNPNVTLDPDGLMPTIABERIAALKKNKVTDSAPSAPANATVAI-----N 710

QY 711 IRPVAKPKSLPK-----ASTSSQP-----TTHPIGAANIKP-----TTSGSIVAPL 753

DB 688 MEEAVYKMKAPQTFKRAIAAQTEQEAHESLTNNP--SVDISPIKNYITDSSQINAAI 745

QY 754 SPVGNKSTSEISLPSAQSSSSSTTSLNLOKSTFLYRADNRSEEMQSKPEGFKAWTP 813

DB 746 RE--NRITPAV---ESLDATSLSSLODRQMRVTVRYVMYVDNS-----TP 784

QY 814 LDTKWARQFASIFIGQKDTSNLPKETVKNISTWGAKPRLKDLNSYIKYTKDKSTVWVSTA 873

DB 785 SPWHSPOEGNSINVGDIVSDNAYLSTSAH-----RGFLNFVHKKETSETRYVQWA 834

QY 874 INTRAGGSSGAPLHKIDMDIYEPAID-----GOKLNPDPGRYKMWVPSILLDTPO 925

DB 835 PLTNAGVNVPAASMYNNAGEQVFMDLNDRSKSLAEKLLKRVSGSQSGQAEIILLPRETQ 894

QY 926 IETSSIIALNH 936

DB 895 FE---VVSMKH 902

RESULT 7

ID ADR18626 standard; protein; 949 AA.

XX ADR18626;

AC 12-AUG-2004 (first entry)

DE Photorhabdus luminescens W-14 tccC4 protein SEQ ID NO:14.

XX tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

KW orally active insect toxin; insect toxin; tccC4.

Query Match	Best Local Similarity	Score	DB 8;	Length	949;
Matches 583; Conservative 100; Mismatches 207; Indels 67; Gaps 17;					
Qy 1	MKNIDPKLVQKTPVSVYDNRGLIIRNIDPHRTTANGDPDTRIRHQYDIGHLNQSIDP 60				
Db 1	MKNIDPKLVQHTFTVNVYDNRGLIIRNIDPHRDVAGDTRIRHQYDTRGHSQSIDP 60				
Qy 61	RLYEAKQTNTIKPNFLQVLDLGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120				
Db 61	RLYDAKQTNNSTPNFLQVNLGDTLRTESVDAGRTVALNDIEGRQVLIVTATGAIQTR 120				
Qy 121	QYETSSLPGLLSVARTPEEKTSRITELRIWAGTAEKDHNLAGQCVRHVDYAGVRL 180				
Db 121	QYEANTLPGLLSVRSQAPCEQTPRVTETFIWAGTQAEKDHNLAGQCVRHVDYAGVQL 180				
Qy 181	ESLSLGTVLSSQLLIDTQENWTDGNETVQWNLADDIYTLTSTFDATGALLTQIDA 240				
Db 181	ESLSLNTILSQRQLLAGQADWDGNETLWQKLNSYTTTQSTFDATGALLTQIDA 240				
Qy 241	KGNIORLAYDVAGQLNGSLTLKGQTEQVVIKSLTVSAAGQKLRHEHNDVITEVSYPE 300				
Db 241	KGNMQRLAYNVAGQLGSLTLKNSQEQVIVKSLTVSAAGQKLRHEHNGVITEVSYPE 300				
Qy 301	TQRLIGIKTRPSDTKVLQDLRYEYDPCGNVLSIRNDAEATPFWNKQVMENTYDLSL 360				
Db 301	TLRLIGITRRQSDSKVLQDLRYEHDHPVGNIIISVRNDAEATPFWNRKIVPENTYDLSL 360				
Qy 361	YQLISATGEMANIGQSHQFPSP--ALPDSNNYTNVTRTYTDGRGNLTKEIHSPPAT 418				
Db 361	YQLISATGEMANIGQSNQSPPIPLPTDENSNTYTRSYNDRGNLQVIRHSSPAA 420				
Qy 419	QNNYTTNITVSNRRAVLSTLTEDPAQVDALFDAGGHQNTLIISQNLNWNTRGELQVIT 478				

CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C  
 CC is encoded by a naturally occurring gene or has an amino acid sequence  
 CC that differs from the product encoded by a naturally occurring gene only  
 CC by truncation or by conservative amino acid changes. Protein A is a 230-  
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic  
 CC species, has stand alone insecticidal activity, and has an amino acid  
 CC sequence at least 40% identical to a sequence selected from xptAlwi,  
 CC xptA2w1, TcdA, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin  
 CC complex potentiator having an amino acid sequence at least 40% identical  
 CC to a sequence selected from TcdB1, TcdB2, TcdB3, TcdC, xptC1w1, xptB1ab,  
 CC xptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator  
 CC having an amino acid sequence at least 35% identical to a sequence  
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, xptC1w1, xptC1ab, PptC1  
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic  
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein  
 CC C. The method is useful for pest control. The present sequence represents  
 CC Photobacterium luminescens TccC4 toxin.  
 XX  
 SQ Sequence 949 AA;

Query Match 57.0%; Score 2850.5; DB 8; Length 949;  
 Best Local Similarity 60.9%; Pred. NO. 3.7e-189;  
 Matches 583; Conservative 100; Mismatches 207; Indels 67; Gaps 17;  
 QY 1 MKNIDPKLYQKTPVSVYDNRGLIRNIDFRTTANGDPDTRITRHOYDIHGHILNQSIDP 60  
 DB 1 MKNIDPKLYQKTPVSVYDNRGLIRNIDFRTTANGDPDTRITRHOYDIHGHILNQSIDP 60  
 QY 61 RLYBAKOTNNIKKNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
 DB 61 RLYBAKOTNNIKKNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
 QY 121 QYETSSLPGRLLSVAEPPEKTSRITERLIWAGNTEAKDHNLGACVRRHYDAGVTRL 180  
 DB 121 QYETSSLPGRLLSVAEPPEKTSRITERLIWAGNTEAKDHNLGACVRRHYDAGVTRL 180  
 QY 181 ELSLSITGVLSQSOLLIDTOEANNVTGNETVWQNLADDIYTLTTFDAGALLTQIDA 240  
 DB 181 ELSLSITGVLSQSOLLIDTOEANNVTGNETVWQNLADDIYTLTTFDAGALLTQIDA 240  
 QY 241 KGNITQRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAAQKJLREHGNVDVITYSEPE 300  
 DB 241 KGNITQRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAAQKJLREHGNVDVITYSEPE 300  
 QY 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISTRNDAEATRFWNNKQMPENTYDLSL 360  
 DB 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISTRNDAEATRFWNNKQMPENTYDLSL 360  
 QY 361 YOLISATGREMANIGQSQSHQSPSP--ALPDSNNVTYNTYTRTYDRCGNLTKIOHSSPAT 418  
 DB 361 YOLISATGREMANIGQSQSHQSPSP--ALPDSNNVTYNTYTRTYDRCGNLTKIOHSSPAT 418  
 QY 419 QNNYTTNITVNSRNRVLSLTEDPAQVDALFDAGGHQNTLISQNLNMTNTRGELQOVT 478  
 DB 419 QNNYTTNITVNSRNRVLSLTEDPAQVDALFDAGGHQNTLISQNLNMTNTRGELQOVT 478  
 QY 479 LVKDKKGDREWRYSGDGRMLKINEQASNNACQRTVYLPNLELRTQNTATTE 538  
 DB 479 LVKDKKGDREWRYSGDGRMLKINEQASNNACQRTVYLPNLELRTQNTATTE 538  
 QY 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLDLSGQIISBEY 598  
 DB 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLDLSGQIISBEY 598  
 QY 599 YPYGCTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQPWIGRWLSSDPAGTIDG 658  
 DB 599 YPYGCTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQPWIGRWLSSDPAGTIDG 658  
 QY 659 LNLVNMVNNPVTLLDPGLMPTTAERTAAUKN-----KVTDSPAPSNATVAIN 710  
 DB 659 LNLVNMVNNPVTLLDPGLMPTTAERTAAUKN-----KVTDSPAPSNATVAIN 710  
 QY 711 IR-----APKPSL--PKASTSSQPTTHPIGAANK--PTTSGSIVAPLSPV 756

Db 711 DEFHYGPDVKKLLEKKPGLNVEELFRGSENGVSTLTKDLPISCISNTSVTLDL 770  
 QY 757 GNKSTSIISLPESAQSSSSTSTNLOKKSFTLYRANRSEEMQSK-FPEGFKAFTPLD 815  
 Db 771 YNKHETK-PPPYENEATVGADLVG-IMSVFEGNKSIGNASDEDLKEEHLPLGKSTMDKTD 828  
 QY 816 TKMARQPASIFIGQKDTSNL-----PKETVKNISTWGAQPKLKLDSNY 858  
 Db 829 LPDLQK--GLMIAEKIKSGKAGYFHFAGAAIAVVGEDKKVAAISILTDLSEPK-RDEGEY 885  
 QY 859 IKYTKDKSTVWVSTAINTEAGGQSSGAPLHKIDMDLYEFA-IDGQKLNPLPEGRTKN 914  
 Db 886 LQSTRKVSAMPI-TNVNEPRG-----HDYPPSKYSIGLVTAERQPVISKRRAN 933  
 RESULT 9  
 ABM70226  
 ID ABM70226 standard; protein; 939 AA.  
 XX  
 AC ABM70226;  
 XX  
 DT 20-NOV-2003 (first entry)  
 XX  
 DE Photobacterium luminescens protein sequence #3323.  
 XX  
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX  
 OS Photobacterium luminescens.  
 XX  
 FN WO200294867-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 07-FEB-2002; 2002WO-IB003040.  
 XX  
 PR 07-FEB-2001; 2001FR-00001659.  
 XX  
 PA (INSP ) INST PASTEUR.  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX  
 WIPI; 2003-148459/14.  
 XX  
 Genomic sequence of Photobacterium luminescens and encoded polypeptides,  
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 Claim 2; SEQ ID NO 3323; 1205pp; French.  
 The invention relates to the isolation of genes and their encoded  
 proteins from Photobacterium luminescens. The isolated sequences are  
 sources of probes and primers for detecting the genome of P. luminescens  
 and related species; to study polymorphisms; for gene analysis and for  
 detection/amplification of the genes. Antibodies (Ab) raised against the  
 polypeptides encoded by the genes are used for detection/identification  
 of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 carry a gene-containing vector are used to select compounds that  
 modulate, regulate, induce or inhibit expression of the genes in plants,  
 animals or microorganisms other than P. luminescens and are able to alter  
 response or sensitivity to toxins and antibiotics produced by P.  
 luminescens. Cells transformed to express the genes are useful for  
 recombinant production of the proteins, particularly toxins and  
 antibacterials useful as insecticides, bactericides and fungicides. The  
 genes, proteins, vectors containing the genes and Ab are also useful  
 therapeutically (to treat microbial infection by bacteria or fungi that  
 are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 biopesticides. Other uses of the genes and the proteins are as virulence  
 factors and for identifying targets of human diseases for which P.

CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX

SQ Sequence 939 AA;

Query Match 56.6%; Score 2833.5; DB 6; Length 939;  
 Best Local Similarity 58.8%; Pred. No. 5.5e-188;  
 Matches 584; Conservative 107; Mismatches 188; Indels 115; Gaps 18;

QY 1 MKNIDPKLYQKPTVTSVVDNGLIRNIDFRTTANGDPDTRITRHOYDIHGHNLQSIDP 60  
 DB 1 MKNIDPKLYHHTPTVSVVDNGLAIRNISFRATAEANTDPRITRHOYNAGYLQSIDP 60

QY 61 RLYEAKQNTNTIKPNFLWOYDLTNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
 DB 61 RLYDAKQTNNAVOPNFVHNLGTNLRTESVDAGRTITLNDIEGRPLVLTISATGVRQNH 120

QY 121 QYETSSLPGRLLSVAEQT-PEKTSRITERLIWAGTEAKDHNLGACQVHYDTAGVTR 179  
 DB 121 LYEDNTLFGRLLAITEQAQTEKT---TERLIWAGNTPQKEYNLVGQCTRHYDTAGLAQ 177

QY 180 LESLSLTGTVLSQSSOLLIDTQEANWNTGNETVQNMMLADDIYTTLTSTFDATGALLTQTD 239  
 DB 178 LNSLALTGAVLSQSQPLVDNQADWNTGDSQMWQKLSVDYVYTONKTDATGALLTQTD 237

QY 240 AKGNIQRLAYDVAGQLNGSWLTLKGQTEQVVIKSLITYSAAGQKLEEHGNDVITEYSYEP 299  
 DB 238 AKGNIQRAYDVAGQLKGSWLTLLKGQTEQVVIKSLITYSAAGQKLEEHGNGIITEYSYEP 297

QY 300 ETORLIGIKTRRPSRTKVLQDLRYEYDPVGNVISIRNDAEATRFWNOKVMPENTYTVDS 359  
 DB 298 ETORLIGITRRPDAKVLQDLRYEYDPVGNVINIRNDAEATRFWRNOKVIPENSYTVDS 357

QY 360 LYQLISATGREMANIGQSHQPPSPALPSDNNYTNRTTYTDYRGGNLTQIQRSSPATQ 419  
 DB 358 LYQLISATGREMANIGQNNPLPSPALPADNNYTNRTTYTDYRGGNLMKIQSSPATQ 417

QY 420 NNYTNTITVSNRRAVLSTLTEDPAQVDALFDAGHQNTLISGQNLNWNTRGELQQVTL 479  
 DB 418 NNYTNTITVSNRRAVLSTLTEDPTQVDALFDGSHQNTLISGQVLWTFRGELKQV-- 475

QY 480 VKRDGANDREWYTSYSGDRMLKINEQQAASNAQTORVYVYLPNLELRITONSTATTED 539  
 DB 476 ---NSSAGNEWYHSDNGTRQLKVNQEQTONIAQQORVYVYLPLELRITQHGSTTTEY 530

QY 540 LQVITVGEAGRAQVRLHWESKPEDINNQLRYSYDNLIGSSOLELSEGOIISERYV 599  
 DB 531 LQVITLKGAGRAQVRLHWESKPEDINNQLRYSYDNLIGSSOLELSEGOIISERYV 590

QY 600 PYGGTALWAARNQTEASYKTRYSKGERDATGLYYGYRYQYQVWIGRWLSDPPAGTIDGL 659  
 DB 591 PFGGTALWAARNQTEASYKTRYSKGERDATGLYYGYRYQYQVWIGRWLSDPPAGTIDGL 650

QY 660 NLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSPANATNVAI-----NIR 712  
 DB 651 NLYRMVRNPNVTFQVQGL-----SPANRTEAIIKQGSFTGME 689

QY 713 PPVAPKPSLPKA-----STSSQPTTHPIGAAN-----IKPTTSGSSIVAPLSVGN 758  
 DB 690 EAVYKMAKQPTFKQRAIATQTEAHQOLLTNPGVDTSPIKOYVTDSSQI-----N 742

QY 759 KSTSE---ISLPESAQSSSSSTTSNLQKSTFLYRADRNSPEEQSKFPFGKAWTPLD 815  
 DB 743 TAIRENRITISVKDLSDLSALQDQIRVYRVWYIDN-----SK-----PSP 786

QY 816 TKMARQFASIFTGQKDTGNLPKETVKNTSTWGAQPKLKDLSNYIKYTKDKSTVWVSTAIN 875  
 DB 787 WHSPQEGNSINVGDIVSDNAVYLSGAH-----RGFLNFVHKETSETRYVKMAFL 836

QY 876 TEAG-----GQSSGNPLHKIDM-DLYBEAIDGQKL--NPLPEGRTKNVPSSLTDT 923  
 DB 837 TNTGVNVAASKYNNENAEKTFKMDLDDSWKSFVEKIKIRANGPAGQAEILFPR---ET 893

QY 924 POIETSSIIALNHGVPVNDAEISFLTTTIPLNKVKP 957  
 DB 894 P-----FEVVSRRHQGRD-----TYVLLQDIKP 916

RESULT 10

ID ABM70214 standard; protein; 960 AA.

XX AC ABM70214;

XX 20-NOV-2003 (first entry)

XX Photorhabdus luminescens protein sequence #3311.

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.

XX Photorhabdus luminescens.

XX WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX Claim 2; SEQ ID NO 3311; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded  
 proteins from Photorhabdus luminescens. The isolated sequences are  
 sources of probes and primers for detecting the genome of P. luminescens  
 and related species; to study polymorphisms; for gene analysis and for  
 detection/amplification of the genes. Antibodies (Ab) raised against the  
 polypeptides encoded by the genes are used for detection/identification  
 of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 carry a gene-containing vector are used to select compounds that  
 modulate, regulate, induce or inhibit expression of the genes in plants,  
 animals or microorganisms other than P. luminescens and are able to alter  
 response or sensitivity to toxins and antibiotics produced by P.  
 luminescens. Cells transformed to express the genes are useful for  
 recombinant production of the proteins, particularly toxins and  
 antibacterials useful as insecticides, bactericides and fungicides. The  
 genes, proteins, vectors containing the genes and Ab are also useful  
 therapeutically (to treat microbial infection by bacteria or fungi that  
 are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 biopesticides. Other uses of the genes and the proteins are as virulence  
 factors and for identifying targets of human diseases for which P.  
 luminescens is a model (particularly plague and whooping cough). This  
 sequence represents one of the isolated P. luminescens proteins

XX Sequence 960 AA;

Query Match 56.4%; Score 2821.5; DB 6; Length 960;

Best Local Similarity 67.0%; Pred. No. 3.9e-187;

Matches 550; Conservative 74; Mismatches 132; Indels 65; Gaps 7;

QY 1 MKNIDPKLYQKPTVTSVVDNGLIRNIDFRTTANGDPDTRITRHOYDIHGHNLQSIDP 60



Db 1 MKNIDPKLYOHTPTISVYDNRGLTIRNIDFHSVAGGDTDTIRTRHQVDVGRHLSQSIDP 60  
 Qy 61 RLYEAKQTNNTIKPNFLWQYDLTGNDPLCTESIDAGRTVTLNDIEGRPLLVTATGVQTR 120  
 Db 61 RLYDAKQDLSINFLWQHLNLTGDTLRTESADAGRTVALNDIEGRQVLIVTATNAIQTR 120  
 Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLAGOCVRYHDYTAGVTRL 180  
 Db 121 QYEAUTLPGRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLAGOCVRYHDYTAGVTRL 180  
 Qy 181 ELSLGTGTVLSQSOLLIDTQEAANNWTDNETVQNMALADDIYTLSTFDTAGALLTQTD 240  
 Db 181 KSLSTENILSQRLLADGQADWTGNNDPLWQTKLSSEIYTTQSTFDTAGALLTQTD 240  
 Qy 241 KGNITQRLAYDVAGQNGSWLTGKQTEQVIKTSITYSAAGOKLREHNGNDVITEYSYEP 300  
 Db 241 KSNITQRLAYDVAGQNGSWLTGKQTEQVIKTSITYSAAGOKLREHNGNDVITEYSYEP 300  
 Qy 301 TQRLIGIKTRPPSDTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKWMPENTYTYD 360  
 Db 301 TRLIGITTRPPSDTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKWMPENTYTYD 360  
 Qy 361 YQLISATGREMANIQGSHQPPSP-ALPDSNNTYNTYTRTYDRGNGNLTKIQHSSPATQ 419  
 Db 361 YQLISATGREMANIQGSHQPPSP-ALPDSNNTYNTYTRTYDRGNGNLTKIQHSSPATQ 420  
 Qy 420 NNYTNTITVNSRNVAVLSTEDPAQVDALFDAGGHQNTLISQNLNWNTRGLOQVTL 479  
 Db 421 NNYTNTITVNSRNVAVLSTEDPAQVDALFDAGGHQNTLISQNLNWNTRGLOQVTL 480  
 Qy 480 VQRDKGANDREWYSGDGRMLKINEQOASNAQOTVRYLNLLELRLTONSTATED 539  
 Db 481 VNREN--TSDEWYRYGNDGRMLKINEQOASNAQOTVRYLNLLELRLTONSTATED 538  
 Qy 540 LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 599  
 Db 539 LHVITVGAAGHAQVRVLHWATTPAGISNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 598  
 Qy 600 PYGTALWAARNQTEASYKTRISGKERDAGLYYGYRYYPQWIGLWLSDDPAGTIDGL 659  
 Db 599 PFGGTALWAARNQTEASYKTRISGKERDAGLYYGYRYYPQWIGLWLSDDPAGTIDGL 658  
 Qy 660 NLYRWVRNPTLLDPLGMLPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 719  
 Db 659 NLYRWVRNPTLLDPLGMLPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 686  
 Qy 720 SLPKASTSSQ-----PTTH---PIGAANIKPTT-----SGSSIVAPLSPVG 757  
 Db 687 PDSEGESESDSICGYRALRSDHPSIHGLRPPPGANLINISAYAHVRAGTSKVKSSWIS 746  
 Qy 758 NKSTSEISLPSAQSSTSTSTNLQKKSFTLYRADNRSFE 798  
 Db 747 YSRSLKVAASWAASGSGRVVKFRICK-----NNRSFD 779

RESULT 11  
 ABM69154  
 ID ABM69154 standard; protein; 966 AA.  
 AC ABM69154;  
 AC ABM69154;  
 DT 20-NOV-2003 (first entry)  
 XX Photorhabdus luminescens protein sequence #2251.  
 XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX Photorhabdus luminescens.  
 OS Photorhabdus luminescens.  
 XX WO200294867-A2.

XX PD 28-NOV-2002.  
 XX PF 07-FEB-2002; 2002WO-IB003040.  
 XX PR 07-FEB-2001; 2001FR-00001659.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 XX PI Buchrieser C;  
 XX DR WPI; 2003-148459/14.  
 XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX PS Claim 2; SEQ ID NO 2251; 1205pp; French.  
 XX CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX SQ Sequence 966 AA;  
 Query Match 55.3%; Score 2768.5; DB 6; Length 966;  
 Best Local Similarity 64.0%; Pred. No. 1.9e-183;  
 Matches 541; Conservative 96; Mismatches 165; Indels 43; Gaps 5;  
 Qy 1 MKNIDPKLYOHTPTISVYDNRGLTIRNIDFHSVAGGDTDTIRTRHQVDVGRHLSQSIDP 60  
 Db 1 MNFPDKLYOHTPTITVHDNRGLDVREIHYHRAEKNSDIRTRHQVDVGRHLSQSIDP 60  
 Qy 61 RLYEAKQTNNTIKPNFLWQYDLTGNDPLCTESIDAGRTVTLNDIEGRPLLVTATGVQTR 120  
 Db 61 RLYDAWQKSSVKNFLWQYDLTGSLVLTESADAGNTVTLHDENRLVLTATGVKQTR 120  
 Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLAGOCVRYHDYTAGVTRL 180  
 Db 121 QYEDNSLPGRLLSVTEQEPDERASHITERLWAGNTPAEKQNLAGOVVHYHDYTAGVTRL 180  
 Qy 181 ELSLGTGTVLSQSOLLIDTQEAANNWTDNETVQNMALADDIYTLSTFDTAGALLTQTD 240  
 Db 181 NHLSTGAALSQSRQLLVQGEADWFGDDKKIWDKKLSDNYVTYTKNTDAGTLLTQTD 240  
 Qy 241 KGNITQRLAYDVAGQNGSWLTGKQTEQVIKTSITYSAAGOKLREHNGNDVITEYSYEP 300  
 Db 241 KGNIOHLAYNVAGQNGSWLTGKQTEQVIKTSITYSAAGOKLREHNGNDVITEYSYEP 300  
 Qy 301 TQRLIGIKTRPPSDTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKWMPENTYTYD 360  
 Db 301 TQRLMGITTRRSTDSKALQDLRYEYDPVGNVISIRNDAEATFRWHNOKWMPENTYTYD 360  
 Qy 361 YQLISATGREMANIQGSHQPPSPALPDSNNTYNTYTRTYDRGNGNLTKIQHSSPATQ 420



Db 361 YQLISATGRENANTGQONLPESTALPSDNNNTYTYAYDRGNLTQIRHSPASQ 420  
 Qy 421 NYTNTITVSNRNLAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNNTTREGELQOQVTLV 480  
 Db 421 NYTNTITVSNRNLAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNNTTREGELQOQVTLV 480  
 Qy 481 KRDK--GANDREWRYSGDGRMLKINEQQAASNNAAQORVYTYLPNLELRLTONSTATTE 538  
 Db 481 KRDLNTRDSDIEWRYGNDGWRLLKVEQEQMQLQORRVVYLSGLELRTTQNGDITKE 540  
 Qy 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNOLRYSDNLISSQLELDSGQIISREY 598  
 Db 541 ELQIITVGEAGRAQVRVLHWSGKPEINNNQVRYSYGNLIGSSQLELDSGQIISREY 600  
 Qy 599 YPYGCTALWAARNQTEASYKTYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658  
 Db 601 YPYGCTALWAARNQTEASYKTYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAGTIDG 660  
 Qy 659 LNLRYMVRNPNVTLDPDGLMPTI-----AER-----TAALKK- 691  
 Db 661 LNLRYMVRNPNVTLDPDGLMPTI-----AER-----TAALKK- 720  
 Qy 692 -----NKVTDSPANATNVAINIRPPVAPKPSLPKASTSSQPTTHPIGAANIKPTT 744  
 Db 721 PAATSWGNKVSQDKTQTLISASAEYGNLKHKKQMEYNNKKLNKSTSPDGKFKYAKTKLODQA 780  
 Qy 745 SGSSIVAPLSPVGNKSTSEISLPSAQSSSTSTTNLOKKSFTLYRADNRSFEEMOSKF 804  
 Db 781 AHAGIAFNHPISONE-----KEGWTFFNYKDFVNLKDKKSLGKGLKPGVELLPDN 833  
 Qy 805 PEGFK 809  
 Db 834 PEHYK 838

RESULT 12  
 ABM69055  
 ID ABM69055 standard; protein; 1044 AA.  
 XX AC ABM69055;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Photothabidus luminescens protein sequence #2152.  
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough.  
 XX OS Photothabidus luminescens.  
 XX FN W0200294867-A2.  
 XX PD 28-NOV-2002.  
 XX PP 07-FEB-2002; 2002WO-IB003040.  
 XX PR 07-FEB-2001; 2001FR-00001659.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;  
 XX WPI; 2003-148459/14.  
 XX Genomic sequence of Photothabidus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX Claim 2; SEQ ID NO 2152; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded  
 CC proteins from Photothabidus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and  
 CC antibacterials useful as insecticides, bactericides and fungicides. The  
 CC genes, proteins, vectors containing the genes and Ab are also useful  
 CC therapeutically (to treat microbial infection by bacteria or fungi that  
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
 CC biopesticides. Other uses of the genes and the proteins are as virulence  
 CC factors and for identifying targets of human diseases for which P.  
 CC luminescens is a model (particularly plague and whooping cough). This  
 CC sequence represents one of the isolated P. luminescens proteins  
 XX

Seq Sequence 1044 AA;  
 Query Match 54.8%; Score 2744; DB 6; Length 1044;  
 Best Local Similarity 74.7%; Pred. No. 1.1e-181;  
 Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3;

Qy 1 MKNTDPKLYQKTPVSYVDNREGLIRNIDFRTTANGDPDTRITRHOYDHLHNSQIDP 60  
 Db 1 MSTPTALYTQTPVSYVDNREGLIRNIDFRTTANGDPDTRITRHOYDHLHNSQIDP 60  
 Qy 61 RLVEAKQNTNNTIKPNFLWQYDLTGNCPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120  
 Db 61 RLVDKQADNSVKPNFVWQYDLHAGLHETESVDAGRTVALNDIEGRPVMTWATGVRQTR 120  
 Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLGAGQCVRYHYDTAGVTRL 180  
 Db 121 RYEGNTLPGRLLSVSEQVDFQETSQVTERFWAENTTAKEYNLSGLCIRHYDTAGVTRL 180  
 Qy 181 ELSLITGTVLSQSSQLLIDTQENANTGNETWQNLADDIYTLTSTFDATGALLTQD 240  
 Db 181 MSQSLAGAILSQSHQWLABGQEAANNVSGDDETFWQMLASEVYTTQSTTNAIGALLTQD 240  
 Qy 241 KGNTRLAYDVAGOLNGSWLTKGTEQVIKSLITYSAGOKLREHNGDVITEYSYRPE 300  
 Db 241 KGNTRLAYDVAGOLNGSWLTKGTEQVIKSLITYSAGOKLREHNGDVITEYSYRPE 300  
 Qy 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHNQKVPENTY 356  
 Db 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHNQKVPENTY 356  
 Qy 357 YDSLYQLISATGRENANTGQONLPESTALPSDNNNTYTYAYDRGNLTQIRHSPASQ 414  
 Db 361 YDSLYQLISATGRENANTGQONLPESTALPSDNNNTYTYAYDRGNLTQIRHSPASQ 420  
 Qy 415 SPATONNTYTNITVSNRNLAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNNTTREGEL 474  
 Db 421 SPATONNTYTNITVSNRNLAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNNTTREGEL 480  
 Qy 475 QQVTLVKRDKGANDREWRYSGDGRMLKINEQQAASNNAAQORVYTYLPNLELRLTONST 534  
 Db 481 QRVTPVSRN-----SSDSEWRYSSDGVRLLYKVEQEQMQLQORRVVYLSGLELRTTQNGD 538  
 Qy 535 ATTSDELQVITVGEAGRAQVRVLHWSGKPEDIDNNOLRYSDNLISSQLELDSGQIIS 594  
 Db 539 KTTQNLQVITVGEAGRAQVRVLHWSGKPEDIDNNOLRYSDNLISSQLELDSGQIIS 598  
 Qy 595 EEEYYPYGGTALWAARNQTEASYKTYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAG 654  
 Db 599 QEEYYPYGGTALWAARNQTEASYKTYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAG 658

```
Qy 655 TIDGLNLYRMVRNPNVTLLDPDGLMPT 681
|:|||||:|||||:|||||:|||||:
Db 659 TVDGLNLYRMVRNPNVTLLDPDGLAPS 685

RESULT 13
AAW18306
ID AAW18306 standard; protein; 1043 AA.
XX
AC AAW18306;
XX
XX 17-OCT-2003 (revised)
DT 30-JAN-1998 (first entry)
XX
XX Photorhabdus luminescens insect toxin TccC.
XX
XX Insecticide; insect; toxin; pest control; biological control;
KW Photorhabdus luminescens; TccC; Southern corn rootworm;
KW Colorado potato beetle; Western corn rootworm; meal worm; boll weevil;
KW turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper;
KW codling moth; corn earworm; European corn borer; tobacco hornworm;
KW tobacco budworm; Lepidoptera; Hymenoptera; Diptera, Dictyoptera; Acarina;
KW Homoptera.
XX
XX Photorhabdus luminescens; strain W-14 (ATCC 55397).
XX
XX WO9717432-A1.
XX
XX 15-MAY-1997.
XX
XX 06-NOV-1996; 96WO-US018003.
XX
XX 06-NOV-1995; 95US-0007255P.
PR 28-FEB-1996; 96US-00608423.
PR 28-AUG-1996; 96US-00705484.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
XX Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;
PI French-Consant RH, Rocheleau RA, Blackburn MB, Hey TD, Merlo DJ;
PI Orr GL, Roberts JL, Strickland JA, Guo L, Ciche T;
XX
XX WPI; 1997-281022/25.
DR N-PSDB; AAT68850.
XX
XX Photorhabdus sp. insecticidal protein toxins and DNA encoding them - can
PT be genetically engineered into insect larvae food and plants for insect
PT control.
XX
XX Claim 34; Page 249-53; 276pp; English.
XX
XX This polypeptide comprises the TccC insecticidal toxin protein of
CC Photorhabdus luminescens W-14. Its amino acid sequence was deduced from a
CC genomic DNA clone (AAT68850). Claimed toxins of P. luminescens (see
CC AAW17871, AAW17884-89, AAW17899-900, AAW18301-06) can be produced by
CC recombinant DNA methods and applied to, or genetically engineered into,
CC insect larvae food and plants for insect control. The toxins are
CC particularly effective against Southern corn rootworm, Colorado potato
CC beetle, Western corn rootworm, meal worm, boll weevil and turf grub
CC (Coleoptera), beet armyworm, black cutworm, cabbage looper, codling moth,
CC corn earworm, European corn borer, tobacco hornworm and tobacco budworm
CC (Lepidoptera), and are also active against insects of the orders
CC Hymenoptera, Diptera, Dictyoptera, Acarina and Homoptera. (All claimed).
CC
XX (Updated on 17-Oct-2003 to standardise OS field)
XX
XX Sequence 1043 AA;
SQ

Query Match 54.7%; Score 2740; DB 2; Length 1043;
Best Local Similarity 74.7%; Pred. No. 2.1e-181;
Matches 513; Conservative 88; Mismatches 88; Indels 8; Gaps 3;

Qy 1 MKNIDPKLYOKTPTVSVLDNRGLSIRDIGFHRIVIGGDTDTTRVTRHQVDARGHLNYSIDP 60
|:|||||:|||||:|||||:|||||:
Db 61 RLYEAKOTNNTIKENFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIOQR 120
|:|||||:|||||:|||||:|||||:
Db 61 RLYDAKQADNSVRKFNFWQHDLAGHARTESVDAGRTVALNDIEGRSVMTWATGVQQR 120
|:|||||:|||||:|||||:|||||:
Qy 121 QYETSSLPGRLLSVAEQTPBEKTSRITERLIWAGNTEAEKDHNLAGOCVRHYDTAGVTRL 180
|:|||||:|||||:|||||:|||||:
Db 121 RYEGNTLPGRLLSVSEQVFQESAKVTERFIWAGNTTSEKEYNLGSLCIRHYDTAGVTRL 180
|:|||||:|||||:|||||:|||||:
Qy 181 ELSLGTGTVLSQSSQLLIDTOEANWTGDNETWQWMLADDIYTLSTFDATGALLTQTD 240
|:|||||:|||||:|||||:|||||:
Db 181 MSQSLAGMLSQSHQLLAEGQEAANWSGDDTVMQGLASEVYTTQSTTGAIGALLTQTD 240
|:|||||:|||||:|||||:|||||:
Qy 241 KGNLQRLAYDVAGOLNGSWLTLKGTOVLIKSLTYSAAAGOKLREEHGNDVITYSEYPE 300
|:|||||:|||||:|||||:|||||:
Db 241 KGNLQRLAYDIAGQLKGSWLTVMQSQGVIVKLSUSAAAGHKUREEHGNGVVTYSEYPE 300
|:|||||:|||||:|||||:|||||:
Qy 301 TQRLIGIKTRR---PSDTTKVLQDLYEYDPVGNVISIRINDAEATRFWNNQKVPENTYT 356
|:|||||:|||||:|||||:|||||:
Db 301 TQRLIGITTRAEQSQSGARVLQDLRYKYDPVGNVISIHNDAEATRFWRNQKVEPNRYV 360
|:|||||:|||||:|||||:|||||:
Qy 357 YDSLYQLISATGRMANIGQOSHQPSPAL--PSDNTYNTYTYTYTYDRCGNLTQKHS 414
|:|||||:|||||:|||||:|||||:
Db 361 YDSLYQLMSATGRMANIGQSNQLPSFVVPVPTDDSTYNTYLRTYTYDRCGNLVQIRHS 420
|:|||||:|||||:|||||:|||||:
Qy 415 SPATONNTYNTIYVSNESRAVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNNTRGEL 474
|:|||||:|||||:|||||:|||||:
Db 421 SPATONSYYTDDITVSSRSRAVLSTLTTPTRVDALFDGSGHQKLLIPGQDLNWNIRGEL 480
|:|||||:|||||:|||||:|||||:
Qy 475 QQVTLVKRDKGANDREWYRYSGDGRRLKINEQQASNNNAQTQRTVYLPNLELRLTQNST 534
|:|||||:|||||:|||||:|||||:
Db 481 QRVTVPVSRN--SSDSEWYRYSDDGMELLKVKSEQQTGNSITQVQRTVYLPGLATGVAD 538
|:|||||:|||||:|||||:|||||:
Qy 535 ATTEDLOVITVGEAGRAQVRVLHWESKPEDIDNNQLRYSVDNLIGSSQLELDSEGGIIS 594
|:|||||:|||||:|||||:|||||:
Db 539 KTTEDLOVITVGEAGRAQVRVLHWESGKPTDIDNNQVRSYDNLIGSSQLELDSEGGIIS 598
|:|||||:|||||:|||||:|||||:
Qy 595 EEEYPPYGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYYPQWIGRWLSSDPAG 654
|:|||||:|||||:|||||:|||||:
Db 599 QEEYPPYGGTALWAARNQTEASYKFIYSGKERDATGLYYGYRYYPQWVGRLWSADPAG 658
|:|||||:|||||:|||||:|||||:
Qy 655 TIDGLNLYRMVRNPNVTLLDPDGLMPT 681
|:|||||:|||||:|||||:
Db 659 TVDGLNLYRMVRNPNVTLLDPDGLAPS 685

RESULT 14
AAW56569
ID AAW56569 standard; protein; 1043 AA.
XX
XX
AC AAW56569;
XX
XX 07-AUG-1998 (first entry)
DT
XX
XX Toxin TccC, encoded by the tccC gene from genomic region tcc.
XX
XX Photorhabdus luminescens W-14; nematode; symbiotic; Heterorhabditis; tca;
KW tcd; tcc; insecticidal activity; toxin; Lepidoptera; Coleoptera;
KW Hymenoptera; Diptera; Dictyoptera; Acarina; Homoptera; Southern;
KW Western corn rootworm; Colorado potato beetle; mealworm; boll weevil;
KW turf grub; beetle armyworm; black cutworm; cabbage looper; codling moth;
KW corn earworm; European corn borer; Tobacco hornworm; budworm.
XX
XX Photorhabdus luminescens.
OS
XX
XX WO9808932-A1.
XX
XX 05-MAR-1998.
PD
XX
XX 05-MAY-1997; 97WO-US007657.
PF
XX
XX 28-AUG-1996; 96US-00705484.
PR
```

Sequence 1043 AA;

Query Match		54.7%;	Score 2740;	DB 8;	Length 1043;
Best Local Similarity		74.7%;	Pred. No. 2.1e-181;		
Matches 513;		Conservative 78;	Mismatches 88;	Indels 8;	Gaps 3;
Qy	1	MKNIDPKLYQKTPVSVYDNRGLIIRINIDFHRITTTANGDPDPTIRTRHQYDIHGHLNQSIDP	60		
Db	1	MSPSETLYTQTPVSVLNDNRLSIRLDIGFHRIVIGGDTDRVTRHQYDARGHLNYSIDP	60		
Qy	61	RLYEAKQTNNTIKNFPLWQYDLTGNPLCTESIDAGRTVTTLNDIEGRPLLVTATGVIQTR	120		
Db	61	RLYDAKQADNSVKFNFWQHDLAGHALRTESVDAGRTVALNDIEGRSVMTMNATGVRQTR	120		
Qy	121	QYETSSLPGRLLSVAEQTPPEKTRITERLIWAGNTRAEKXDNLAGOCVRRHYDTAGVTRL	180		
Db	121	RYEGNTLPGRLLSVSEQVFNQESAKVTERFWAGNTTSEKEYNLSGLCIRHYDTAGVTRL	180		
Qy	181	ESLSLTGTVLSQSQLLIDTQEAANWTGDNETVQNMMLADDIYTYTLSTFDATGALLTQTD	240		
Db	181	MSQSLAGAMLQSQSHQLLAEGQEAANSGDDETVWQMLASEVYTTQSTTNAIGALLTQTD	240		
Qy	241	KGNIQRLAYDVAGQNLNSWLTGKQTEQVLIKSLTYSAAQOKLREHGNVDVITEYSYEPE	300		
Db	241	KGNIQRLAYDIAGQLKGSWLTGKQSEQVIVKSLWSAAGKLEHNGVVTTEYSYEPE	300		
Qy	301	TQRLIGIKTRR- --- PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHQKVMPENTYT	356		
Db	301	TQRLIGITTRAEQSGQARVLQDLRYKIDPVGNVISIHNDAEATRFWRQKVEPENRYV	360		
Qy	357	YDSLYQLISATGREMANIGQSHQFPSPAL--PSDNNYTYNTYTYTYDRGGLTKIOHS	414		
Db	361	YDSLYQLMSATGREMANIGQSNQLPSPVIPVPTDDSTYNTYLTYYTVDRGGLVQIRHS	420		
Qy	415	SPATQNNYTTNITVSNRSNRAVLSTLTEDPAQVDALFDAGCHQNTLISGQNLNWNTRGEL	474		
Db	421	SPATQNSYTTDITVSSRSNRAVLSTLTTPTRVDALFDSGCHQKLI PGQNLWNIRGEL	480		
Qy	475	QOVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNACTQRTVYLPNLELTLQNST	534		
Db	481	QRVTPVSRN--SSDSBWRYSSDGMRLLVSEQQTGNSTQVQRTVYLPGLLELTGVAD	538		
Qy	535	ATTEDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIIS	594		
Db	539	KTTEDLQVITVGEAGRAQVRVLHWESGKPTDIDNNQVRYSYDNLIGSSQLELDSEGOILS	598		
Qy	595	EEEYYPYGGTALWAARNQTEASYKTIYRYSKGERDATGLYYGYRYQYPWIGRWLSSDPAG	654		
Db	599	QEEYYPYGGTAIWAARNQTEASYKFIYRYSKGERDATGLYYGYRYQYPWVGRWLSADPAG	658		
Qy	655	TIDGLNLYRMVRNPNVTLLDPDGLMPT	681		
Db	659	TVDGLNLYRMVRNPNVTLLTDHDLAPS	685		

Search completed: February 16, 2006, 21:31:33  
Job time : 100.189 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 33.2035 Seconds  
(without alignments)  
3670.210 Million cell updates/sec

Title: US-10-754-115-45  
Perfect score: 7901.  
Sequence: 1 MNSQDSITELSLPKGGGA.....WFTVNEBNDTAAEVKKVKM 1474

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgm2\_6/ptodata/1/iaa/5-COMB.pep.\*  
2: /cgm2\_6/ptodata/1/iaa/6-COMB.pep.\*  
3: /cgm2\_6/ptodata/1/iaa/H-COMB.pep.\*  
4: /cgm2\_6/ptodata/1/iaa/pCTUS-COMB.pep.\*  
5: /cgm2\_6/ptodata/1/iaa/RE-COMB.pep.\*  
6: /cgm2\_6/ptodata/1/iaa/backfiles.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6042	76.5	1481	2	US-09-251-645-14
2	6031.5	76.3	1476	2	US-09-817-514A-4
3	4581.5	58.0	1485	2	US-08-851-567B-32
4	160.5	2.0	1426	2	US-09-492-709A-340
5	155.5	2.0	658	2	US-09-252-991A-24910
6	155.5	2.0	2315	2	US-09-543-681A-5434
7	154	1.9	4630	2	US-09-091-609-2
8	154	1.9	5215	2	US-09-105-537-2
9	153.5	1.9	1028	2	US-09-543-681A-7181
10	144	1.8	1377	2	US-09-711-164-467
11	143	1.8	798	2	US-09-489-039A-10045
12	142.5	1.8	1183	1	US-08-447-031A-2
13	140	1.8	979	1	US-08-346-455B-38
14	140	1.8	979	2	US-08-977-221-38
15	140	1.8	979	2	US-09-483-821B-70
16	140	1.8	979	4	PCT-US95-06613-38
17	139.5	1.8	646	2	US-09-902-540-10353
18	139.5	1.8	1577	1	US-08-793-824-2
19	139	1.8	2200	2	US-09-796-575-2
20	138	1.7	1529	2	US-09-215-694-1
21	138	1.7	1529	2	US-10-109-310-1
22	137	1.7	2628	1	US-08-570-311-14
23	136.5	1.7	990	1	US-08-232-540-2
24	136.5	1.7	990	1	US-08-428-949A-2
25	136.5	1.7	990	1	US-08-428-948A-2
26	136.5	1.7	990	1	US-08-428-946-2
27	136.5	1.7	990	4	PCT-US95-04656-2

## ALIGNMENTS

### RESULT 1

US-09-251-645-14  
; Sequence 14, Application US/09251645  
; Patent No. 6281413  
; GENERAL INFORMATION:  
; APPLICANT: Kramer, Vance C.  
; APPLICANT: Morgan, Michael K.  
; APPLICANT: Anderson, Arne R.  
; APPLICANT: Hart, Hope  
; APPLICANT: Warren, Gregory W.  
; APPLICANT: Dunn, Martha  
; APPLICANT: Chen, Jeng S.  
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS  
; FILE REFERENCE: CGC1963/A  
; CURRENT APPLICATION NUMBER: US/09/251,645  
; CURRENT FILING DATE: 1999-02-17  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 1481  
; TYPE: PRT  
; ORGANISM: Photorhabdus luminescens  
US-09-251-645-14

Query Match 76.5%; Score 6042; DB 2; Length 1481;  
Best Local Similarity 76.1%; Pred. No. 0;  
Matches 1127; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy	1	MNSQDSITELSLPKGGGALTGMGEALTPTGPGMAALSPLPISACRGVAPFTLVN	60
Db	1	MNSQTSMTLSLPKGGGALTGMGEALTPTGPGMAALSPLPISACRGVAPFTLVN	60
Qy	61	SGAGNSPFLGWCNVMVTRRRTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDS	114
Db	61	SGTNSPFLGWCNVMVTRRRTHFGVPHYDETDTFLGPEGEVLVAALNEAGQADIRSES	120
Qy	115	TLOGINLGAFTVTVGYSRLSHFSRLBYWPKTKTDFWLIVSPDGOVHLLGKSPQAR	174
Db	121	SLOGINLGMTFTVTVGYSRLSHFSRLBYWPKTKTDFWLIVSPDGOVHLLGKSPQAR	180
Qy	175	ISNSQTTQTAQWLEASVSSRGQIYYOYAEADDTCGEADEITHLQATQRYLHVY	234
Db	181	ISNPLNVNQTQWLEASVSSRGQIYYOYAEADDTCGEADEITHLQATQRYLHVY	240
Qy	235	GNRTASELPLGLDGSAPSQADWLFVLPDYDYSRNLKTPPAFTTSGWLCRQDRFSRYE	294
Db	241	GNLTASEVFTLNGDDPLKSGWLFCLVFDYGERKNSLSEMPFPKATSNWLCRQDRFSRYE	300
Qy	295	YGFEIRRLCRQVLMYHHLQALDSKITEHNGPTLVSRILNLYDESAIASTLVFVRVGH	354

Sequence 8, Appli  
Sequence 1, Appli  
Sequence 4294, Ap  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 36, Appli  
Sequence 36, Appli  
Sequence 36, Appli  
Sequence 3, Appli  
Sequence 23805, A  
Sequence 4476, Ap  
Sequence 6274, Ap  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 21474, A  
Sequence 14853, A





```
Qy 172 QARISNPSQTTOTTAOWLLEASVSRGEIOYVQRAEDDTGCEADEITHLQATAQRYLHI 231
Db 180 QACLANPONQOQTAQWLEBETVTPAGEHVSQYRAEAEACDDNEKTAHPNVAQRLVQ 239
Qy 232 VYIGNRTASETLPGLDGSAPSQADWLFYLVDFYDGERSNKLTTPAFST-TGSMLCRODRF 290
Db 240 VNYGNIKPQASLFVLDNAPAPAEBSWLFHLVFDHGERDTSLHTVPTWDAGTAQMSVRPDI 299
Qy 291 SRVEYGEIRTRLCROVLMYHHQLQALDSKITEHNGPTLVSRLLINYNDESALSTLVFR 350
Db 300 SRVEYGEVTRRLCCQVLMFHRTALMAGEASTNDAPELVGRLLILEYDKNASVTLTIR 359
Qy 351 RVGHEQDGNVTLPLPBLAYQDPSRHHQWPMVLANAQRWOLVDLKEGELPGLL 410
Db 360 QLSHESDGRPVTPPLELAWQRFDEKIPWQRFDALDNFSQORYQLVLRGEGLPGL 419
Qy 411 YQKGAWYRSARQLRGISGDAVTWEKMQPLSVIPSLQSNASLVDINGDGLQDWMVITPG 470
Db 420 YQDGAWYKAPQREQDGSNAVYDKIAPLPTLNLQDNASLMDINGDGLQDWMVITAG 479
Qy 471 LRGHVSQRPDGSWTRFPTLNLALPVEYTHPRAQALADLMGAGLSDLVLIGPKSVRLYANTR 530
Db 480 IRGHVSQRPDGSWTRFPTLNLALPVEYTHPRAQALADLMGAGLSDLVLIGPKSVRLYANTR 539
Qy 531 GFAGKDVVQSGDITLTPVPGADPRKLVAFSDVLGSGOAHLEVESATKVTCPNLRGRFRG 590
Db 540 GWRKGDVPGSTGLTLPVTGTDAKLVAFSDVLGSGOQHIVEIKGNVTCWPNLGHGRFRG 599
Qy 591 QPITLPGFSOPATFENPAQVYLADLDGSGPTDLIYVHTNRDLIDPLNKGSGFAPBPTLRF 650
Db 600 QPITLPGFSOPENSFENRFLADLDGSGPTDLIYVHTNRDLIDPLNKGSGFAPBPTLRF 659
Qy 651 PEGLRFPHTCOLQADVOGLVGLASLILSVPHMSPHHRCDLTNWKPWLLNEMNNMNVHH 710
Db 660 PEGVQFQNTCOLQADVOGLVGLASLILSVPHMSPHHRCDLTNWKPWLLNEMNNMNVHH 719
Qy 711 TLYRSSSQWLBDEKAALTTGOTFVCLPFPPLHTLMOTETEDISGNKLVTLRYARGA 770
Db 720 TLYRSSSQWLBDEKQLTVAGKSPACLPFPPLHTLMOTETEDISGNKLVTLRYARGA 779
Qy 771 WDGRERFRGFGVYEQDTHSLQAGNAPERTPALTKNMYATGLPVIDNALSTEYMR-DD 829
Db 780 WDGRERFRGFGCQKQDITTFHSGTAPQAPASLSISWFAWGDEVDSQLATEYMQADT 839
Qy 830 QAFAGSPRTTWQ--DNKDVPLTPEDDNRVFNWRAKQQLRSELYGLDDSTNKVFPY 887
Db 840 QAYSGFTRYTVMDHTNQDQAFETP-NETQRNMLTRALKGQLRLTELYGLDGTDKQTPY 898
Qy 888 TVTEFRSQVRQLQHTDSRYVPLMSSVVESEYHYERTASDPQCSNITLSSDRFGQPLKQ 947
Db 899 TVBESRYQVRSIPVNMKTELSAWTALENSXYHYERTITDPQFSQSLKQHDIFGQSLQS 958
Qy 948 LSVQYPRRQOPAINLYPDTLPDKLLANSYDDQORQLRTYQSSWHHLTNNTVRVLGLPD 1007
Db 959 VDIAPWEREPANVPYPTLPETLFDSSYDDQQQLRLVQKNSWHHLTDGENWRGLPN 1018
Qy 1008 STSDIPTYGAEVNPAGGLNLELLSDKNSLIADKPREYIQQOKTATYDQNTYPTLOPT 1067
Db 1019 AQRDVTYDRSKIPTREGISLEILLKDDGLLADDEKAAVYLGQQQTFYTAGQAEVLEKPT 1078
Qy 1068 RQALIAPTETTVFNOSTLSAFNGSIPSDKLSLTLBQAGYQOQTNVLRPRTGEDKVVAAHG 1127
Db 1079 LQALVAFQETAMDDTSLQAYEGVIEQELNTALTQAGYQOQVAKLFNTRSESPWAAQ 1138
Qy 1128 YTDYGTAAQFWRPQKQNTQLTGKITLIWDAVYCVVQTRDAAGLTTSKAYDWRFLTPVQ 1187
Db 1139 YTDYGDAAQFWRPQKQNTQLTGKITLIWDAVYCVVQTRDAAGLTTSKAYDWRFLTPVQ 1198
Qy 1188 LTDINDNQHILTDALGRPTLRFWGTENGKMTYSSPEKASPSPPSDVNAALTELKPLP 1247
Db 1199 LTDINDNQHIVTLDALGRVTTSRFPWGTGAGQAAGYSN---QPETPPDSVDKALALTALP 1255
```

```
Qy 1248 VAQCOVYAPESWMPVLSQKTFNRL---AEQDWOKLYNARIITEDGRICCTLAYRRWVQSK 1304
Db 1256 VAQCLVAVDSWMPVLSLSLSQSQSEAEALWALRAAHMITEDGKVICALSGKRGTSQON 1315
Qy 1305 AIPQLISLLNNGRPLPHPSLTLTTDRYDHPDQOIROQVVPFSDGFGRLQAAARHEAGMA 1364
Db 1316 LTIQLISLLASIPRLPPHVLGITTDYDSDPOQHQOQTSFSDGFGRLQSSARHESGDA 1375
Qy 1365 RQNEGSGLIINVQ-----HTENRVAVTGRTVDNKGQPIRTYQVFLNDRWVVSND 1418
Db 1376 WQRKEGGLVWDANGVLVSAPTDRWAVSGRTYDDKQQPVRTYQVFLNDRWVVSND 1435
Qy 1419 RQKEAYADTHVYDPIGRIKVTAKGFRRTLTFTWFTVNEDENDTAA 1467
Db 1436 RD-DLFADTHLYDPLGREYKVTAKKYLREKLYTPWFIVSEDENDTAS 1482
```

## RESULT 4

```
US-09-492-709A-340
; Sequence 340, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ESCHERICHIA COLI
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
; US-09-492-709A-340
```

```
Query Match 2.0%; Score 160.5; DB 2; Length 1426;
Best Local Similarity 19.0%; Pred. No. 0.00045;
Matches 255; Conservative 152; Mismatches 446; Indels 487; Gaps 65;
```

```
Qy 406 LPGLLYQDKGAWYRSQAQLRGISGDAVTWEKMQPLSVIPSLQSNASLVDINGDGLQD 465
Db 70 LPFILSRYSYRTKTPAPVGVFGPG---WK--APSDIRLQLRDDGLILNDNGRSIHFE 124
Qy 466 ITGPG-----LRGHSQRPDGSWTRTFPLNALPVEYTHPRAQALADLMGAGLSD 513
Db 125 PLLPGEAVYRSRSMVLVRGGKAAQPDGH-----TLRLWGLALPPD 165
Qy 514 LVLIQPKSVRLYANTRDGFAGK-----KDVVQSGDITLTPVPGADPRKLVAFSDVLG-- 564
Db 166 IRL-----SPHLYLATNS--AQGFNWLGNSEYVPGAEVVLPAFLPPVRLVTGMADRFRT 219
Qy 565 -----SGOAHLEVESATKVTCPNLRGRFGQPI 594
Db 220 LTVRREAAGLAGEITGVTDGAGREFRLVLTQQAAREARTS-----SLSSDSSRPLS 274
Qy 595 LPGF--SQPATEFNP-----AQVYLDLDGSGPTDL-----IVVHTNRDLIDFLNKS 638
Db 275 ASAFPDTLPGETYGPDRGIRLSAVLMLH--DPAYPESLPAAPLVRYTTEAGELLAVYDRS 333
Qy 639 GNGFAEPVTLRFPEGLRFDHTCQLQADVOGLVGLASLILSVPHMSPHHRCDLTNMKPWL 698
Db 334 NTQV-----RAFTYD-----AQHFGRWVAHYACRPEM---RYRYDDTGR---- 370
Qy 699 LNEKNNMNGVHHTLRYRVSRSQFWLDEKAAALTTGQTPVCLYPPPIHTLMOTETEDISGN 758
```



```

Db 371 VVEQLNPAGLSVRYLYEQ-----DRITVDSLNREV-----LHT-----EGGAGL 411
Qy 759 KLVTTLRYARGAWDCRREFRGFGYVEQTDHSLQAQGNAPRTTPALTKNYATGLPVID 818
Db 412 KRVVKELADGS-----VTRSGYDAAGRLTAQTDAAARRTEY--GLNVVS 454
Qy 819 NALS-----TBYWDD--QAPAGSPRFTTWQDNKVDPLTPEDDSDNSRYWFRALKG 867
Db 455 GDITDITTPDGRETFFYNDGNQLTAVVSPDGLSREYDEFGRLVSETSR-----SG 507
Qy 868 QLLSELGLDDSTNKHVPYVTEPRSQVRLQHTSDSPVPLWSSVVSERNVHYERIASD 927
Db 508 ETVR---YRYDDA--HSELPAITTDATGSTRQ-----TWS-----RYGQLLAP 546
Qy 928 PQCSQNIIT--LSSDRFGQPLKQLSVQPRRQPAINLYPDTLFDKLLANSYDQDQRLTLT 986
Db 547 TDCSGYQRYEYDRFGQ-----MTAVHREBEGISLYRYDNRGRLTSVKAQGRETYE 599
Qy 987 Y-----QOSSWHHLTNNTV-----RVGLPDSTRS--DI 1013
Db 600 YNAAGDLTAVITPDGNSRSETQYDAMGKAVSTTQGGLTRSMEDYDAAGRVISLTNENGSHV 659
Qy 1014 FTYGAEN--VPAGGLN-----LELLSPKNSLIA-----1039
Db 660 FSYDALDRLVQGGPDGRTQRYHYDLTKLTQSEDEGLVILWYDESRIHRTWTNGEPA 719
Qy 1040 -----DOKPR-----EYLGQOKT--AY 1054
Db 720 EQWQYDGHGWLTDISHLSEGRVAVHYGYDDKRLTGECQIVENPETGELLWQHETKAY 779
Qy 1055 TD---GQNTTLPQTTRQ-----ALIAFTTTFVNSTLS--AFNGSI 1092
Db 780 NEQGLANRVTPDLSPPVWELTVGSGYLACMKLGTPPLVEYTRDLRHRETIVRSFGSMAGN 839
Qy 1093 PSDKLSLTLEAGYQOYNLTPRTGEDKVAHAGYTDYGTAAQFWRPQ--KQSTQLTGT 1150
Db 840 AAYELTSTYTAGQLQSOHL-----NSLYDRDYGWSDGDLVRLSGPRQTRYGYSATG 894
Qy 1151 KI---TLIWADANYCVVQTRDAAG-----LTSYAKYDWRFLTP 1185
Db 895 RLESVRTLAPDLDIRIPAT--DPAGNRLPDPELHPDSTLTWPDNRIAEDAHYVVRHDEY 953
Qy 1186 VOLTIN-----DNQHLITL-----DALGRPIT 1208
Db 954 GRLEKTDRIIPAGVIRTDDERTHYHDSQHLVETRYIOHGEPLVESRYLYDPLGRMA 1013
Qy 1209 LRFWCTENGKMTGYSS-----PEKASFSPSDVNAIELKPLPVAQCQ--VYAPESWMPVL 1263
Db 1014 KRVWRRER-DLTGMWLSRKPEVTWYWGWDRLTTVQ-----TDTTRIGTVYEPGSGFTPLI 1068
Qy 1264 SOKTFNRLAEQDWQKLYNARIITEGRICTLAYRRWVQSKAIPOLISLLNNGPRLP--PH 1322
Db 1069 RVETENGREKA--QRRSLAETLQEGS-----ENGHVVEPA 1104
Qy 1323 SLTLTDRYHDPEQIQVQVFSDFGRLGLQAARHEAGMARQNE--GSLINVOHT 1380
Db 1105 ELVRLDLRL-----EBEIRADRVSSESLAWLAQCGLTVEQ--LARQVEPYTPARKAHLVHC 1159
Qy 1381 ENR-----WAVYGRTEYD-----NKGQPIRTYQPYFLNDRYVNSDSARQ 1420
Db 1160 DHRGLPLALISBDGNTAWS-----AEYDEWGNQLNENPHHYVYQYRPLFGQOH-----DE 1209
Qy 1421 EKEAYADTH--VVDPI--GREI 1438
Db 1210 ESGLYNRHRYDPLQGRYI 1229

```

RESULT 5  
US-09-252-991A-24910  
; Sequence 24910, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24910  
; LENGTH: 658  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; US-09-252-991A-24910

Query Match 2.0%; Score 155.5; DB 2; Length 658;

Best Local Similarity 20.1%; Pred. No. 0.00031;  
Matches 113; Conservative 61; Mismatches 155; Indels 233; Gaps 31;

```

Qy 356 ODGNVVTLPPLLELAYQDFSPRRHHAHQPMQVLANFNAIQRWQLVDLKGGL--PGLLYQ-- 412
Db 123 REGDNVTLVRNRLAEDTS---IHWGIIILPANMDGVP---GLSFGIAPGGLYEVR 173
Qy 413 ---DKGAWVTRSAORL-----GETGSDA-----VTWEKQPLSVIP 445
Db 174 FKVRQNGTYWTHSHSGLQEQAGVYGALVIDAREPEPSYDRDYVYVLLSDWSEKQRIILA 233
Qy 446 SLQSNAS-----LVDINGDQGLQDWWIT-----GP---GLRGY--- 474
Db 234 KLKQSDYNNFKRTVGDGFIDDSANG---WAATLADRKMAEMKMSFTDLADVSGYTYT 290
Qy 475 ---HSRQPDGSGWT-----RFTPLNAL-----PVEYTT 497
Db 291 YLLNGQPPDGNGWTLFRPGEKRLRFVNASAMSYFVRI PGLKMTVVAADGQHVEPVSVD 350
Qy 498 HPRAQLADLMGAGLSDLVLGPKSVRLY-----ANTRDGPAGKDVQSGDITLTPVPGAD 552
Db 351 ELRIATAETVD-----VIVEPQGERAYTLFAQMDRSRGYARGTLALAEGL--LSAPVPTPD 403
Qy 553 PRKLVAFSVDVLGSOAH-----LVVSATKVTCPNLTG-----585
Db 404 PRPLIGMDDMGMGDGHGAMGHGAATRPASENDSHSKSGMDKMGDHSKMGAMDMNGMDH 463
Qy 586 -----RGRFGQPIITLPGFSQPAE--FNPAQVYVLADLDGSGPTDLYVYHTNRLDIFLNK 637
Db 464 SKMAGMDHSRGMG--AMPQSHPASEDGNP---LVDMTQWTPTPKL-----ADPGLGL 512
Qy 638 SGNG-----PAPVTLRP--PEGLRFHDTCOLQADVOGLGVASLILSVPHMSPHHWCD 690
Db 513 RDNRRRLVTYAD--LRSRFADPDGREGPRTIELHTG-----HMEKFAWSFD 557
Qy 691 ---LTNNKPLWLN-----MNNMGVHHTLVRSSSQFWLDEKAAALTTGTFVCYL 739
Db 558 GIKFSDAEPRLKTYGERLIRVLVNDTWMT-----587
Qy 740 PFIHTLWQTEDEDEISGNKLV 761
Db 588 PIHLHGMW--SLEDE--QGNFLV 607

```

RESULT 6  
US-09-543-681A-5434  
; Sequence 5434, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706

; PRIOR FILING DATE: 1999-04-09  
 ; NUMBER OF SEQ ID NOS: 8344  
 ; SEQ ID NO 5434  
 ; LENGTH: 2315  
 ; TYPE: PRF  
 ; ORGANISM: Proteus mirabilis  
 US-09-543-681A-5434

Query Match 2.0%; Score 155.5; DB 2; Length 2315;  
 Best Local Similarity 20.0%; Pred. No. 0.003;  
 Matches 313; Conservative 179; Mismatches 558; Indels 513; Gaps 81;

Qy	10	TELSLPKGGGAGTGMGALTPGDP-----GMAALSPL-PI-SAGGY 51
Db	662	TEVTVP--GGSAAGDTLTLTKPDGSDTDTVEHTLTADVTAGKADVTIIPADKVTADGNY 719
Qy	52	APAFITLNYSGAGNSPGLGWCNMVTRRTHRGVPHYDSTDFTLGESEVLVADQPR 111
Db	720	--SVTAETIDPAGNTS--GQG-----KPTDFAVDTVAPAPVLKAEDDDGVSVDLPT 767
Qy	112	DESTLOGINLGATTVTGYRSLRSHFSRLEYWQPKTKTDFWLIYSPD-----161
Db	768	D-----ANKGDTVEITFDEKGGKHTVLEKGDNGWTSPTPALI-----PDSNGDKATIPA 818
Qy	162	-----GOVHLLGKSPOARISNPSTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEAD 215
Db	819	DNVKDNEVTGIAKDPGNSDPSTVTSKTDGVADAPVLTIPEVA-----DGYANAD 870
Qy	216	EITHLQA-----TAQRYLHV-----232
Db	871	ELKDGQAQAEVLPAGTVEGAEITLTVTRPKDTTETVTHVTVKDEAAAGKVSVDIPKDAVQ 930
Qy	233	-----YYGRTASETLPG-----LDGSPSQADWLFLYVFDYGRSNMLKTPPA 276
Db	931	NGQNSVDVSLTQGNPAK---PGNKVDFAVDGQIPGSDT-----GDGFTVDTPV 976
Qy	277	FSTTGMCLCRDRFSREYGEFEITRRLCRQVLMVHLLQALDS---KITHEHNGPTLVSRLL 333
Db	977	VTIPEA---ADGVNADELKDGQVQ-----QVTVPGSSAAGDTLTLTKPDGSDTIVEH 1027
Qy	334	ILNYDESAIASTLVFRRVGHGEQGNV-----VTLP-----PLELAYQDFSPRHH 379
Db	1028	TLTADVTAGKADVTIPADKVTADGNTSVTAETIDPAGNTSGQKPTDFAVDTVAPS---1084
Qy	380	HWQP-----MDVLNPNALQWQV--DLKGEGLPGLLYQDKGAWWYSQAQLG 426
Db	1085	--APVLKAEDDGSVSDLPFTDANKGDTVEITFDEKGGKHTVLEKGDNGWTSPTPALIP 1142
Qy	427	EIGSDAVT--WEKQPLSVIPSL-----QSNASLVLDINGDQLDW-VITGPGLR-GYH 475
Db	1143	DSNGDKATIPADNVKDNSEVTGIAKDPGNSDPSVTSTKTDGVADAPVLTIPEVADGYA 1202
Qy	476	S--QRPGSWTRFTPLNALPVEYTHPRAQLADLWGAGLSLVLIGPKSVRLYAN--TRDG 531
Db	1203	NADELKDG-----LQAEVTLPACTIE--GAEITLVTRPDKTTETVTHVTKDE 1249
Qy	532	PAKG-----KQVQSG-----DITLPVGADPRK---LVAFS-----DVLGSGQAH 570
Db	1250	AAAGKVSVDIPKDAVQNGQNSVDVSL--TQGNPNPAKGNKVDFAVDGQIPGTDGSG---1304
Qy	571	VEVSATKVTCPNPLGRG-----RFQPIITLPFSQAPATEFNPAPQVYLADLDGSGPTD 622
Db	1305	-TVTTPVVTIPEATDGVNADELKDGQVQTEVTVPGSSAAG---DTLTLTIKPDGS--TD 1358
Qy	623	LI-----YVHTNRLDIFL-----NKSNG-----GFAEPVTLRPPGRLRPHDTCQIQM- 664
Db	1359	TVEHTLTADVTAGKADVTIPADKVTADGNTSVTAETIDPAGNTSGQKPTDPMVDVTQIP 1418
Qy	665	ADVGLGVASL--ILSVPHMSPHWRCDLTNWKPLLNMNNMNVGHHTLYRYSRSSQFWL 722
Db	1419	GDTGDGVDVDTTPVVTIPEAT-----DGVN-----ADELKGQVQTEVTVPGSGAA---1463
Qy	723	DEKAAALTGTQTPVCYLPPF---IHTLWQTEDEISGNKLVTLTLRYARGAWDGREREFR 779

Db	1464	-----GDTLTLTKPDGSDTDTVEHTLTADVTAGKADVTIIPADKATPDG-----1508
Qy	780	GFGYVEQTDHQLAQGNAPERTPPALTKWYATGLPVI-----DN-ALSTEVWRDQAPAGF 835
Db	1509	--NYSVKAETIDPAGNTSGEGKATDFTVDTVAPSTPVLNAEDNGSVSVELPGD--ANKGD 1564
Qy	836	SPRFTWQDNK-----DVPLTPEDDNRVWFNRALKQGLLR--SELY 875
Db	1565	TVEI-TPEDEKGGKQVTMBKGDNGWTSNPLIPDQGN-----NTAIPSDNVKDNSEVT 1619
Qy	876	GL--DDSTNRKHVPYVTFEPRSVRRRLQHTDSRYPVLWSSVVESSRNHYERIASPQCSQ 933
Db	1620	AIKADPGNSAPATA-----MSKTDVLPVTSISVDTTSDVNDGDKISADASVSGE 1670
Qy	934	ITL-----SSDRFGQPLKQLSVQYPRRQOPAINLPDTL-----PD-KLLANSY 976
Db	1671	ITEVPATIEDKDDTTGL-VYTVALDVYAAQDVTVTI---TLTNDAGHASAPDYSTLAGSQ 1726
Qy	977	DDQORQL-----RLTYQQSSWHHL-----995
Db	1727	HDGKIALHGDGKVTYDGCSTVTVVTPAGSKSVSFIYDPTMEANQNAFNAEGMEKVATI 1786
Qy	996	---TNNTVRVL-----GLPDSTRS-----DIFTYGAEN-----1020
Db	1787	TGTSNNATAVTDVNNAGASATGVIYDGNPISLRNLDDFTLKYSLSSVAENGDFGVTI 1846
Qy	1021	VPAGGLNLELSDKNSLIADDPREYL-----GQOKTAYT-----DGQNTTPLQ 1064
Db	1847	GVDSGKKDKPLTDDYNDTI-----YVGYQDGSETSSYSNLSANSDNGPDGKTGDNQ 1899
Qy	1065	TPTRQALIAFTETTVFNQSTLS-----APNGSIPSDKLSLTLEQ--AGYQQTNYLPFR 1115
Db	1900	SITTVDLGAGDMLVIRGNMLTNRVYAGENDFTMDGNALRSWAG-----SYLFME 1955
Qy	1116	TGEDKVVVAHHGYTDYG-----TAAQFWRPQKOSNTQLTGKITLINDANYCVVVQT 1166
Db	1956	SGNDTVTIKRTGVTNAQOIVLGSSTFTQ--GDANDQNTTSLGLDL-----2002
Qy	1167	RDAAGLTSYKDWRFITPVQ-----LTDINDNQHLLTLDALGRPITLRFWGTENGK-WTG 1221
Db	2003	--GSGTQDKSNMPNEXLSVYQDGSNLSLGN-----NIDAVTDVNTVTIYGSVSGEILGG 2056
Qy	1222	YSS 1224
Db	2057	YGS 2059

RESULT 7  
 US-09-091-609-2  
 ; Sequence 2, Application US/09091609  
 ; Patent No. 6600029  
 ; GENERAL INFORMATION:  
 ; APPLICANT: SHERMAN, DAVID H.  
 ; APPLICANT: WILLIAMS, MARK D.  
 ; APPLICANT: XUE, YONGQUAN  
 ; TITLE OF INVENTION: METABOLIC ENGINEERING OF  
 ; POLYHYDROXYALKANOATE MONOMER SYNTHASES  
 ; FILE REFERENCE: 600.297US2  
 ; CURRENT APPLICATION NUMBER: US/09/091.609  
 ; CURRENT FILING DATE: 1998-06-19  
 ; EARLIER APPLICATION NUMBER: PCT/US96/20119  
 ; EARLIER FILING DATE: 1996-12-18  
 ; EARLIER APPLICATION NUMBER: 60/008,847  
 ; EARLIER FILING DATE: 1995-12-19  
 ; NUMBER OF SEQ ID NOS: 4  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 4630  
 ; TYPE: PRF  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-091-609-2

Query Match 1.9%; Score 154; DB 2; Length 4630;  
Best Local Similarity 21.0%; Pred. No. 0.014;  
Matches 247; Conservative 123; Mismatches 426; Indels 380; Gaps 65;

QY 189 LEASVSSRGE---QIYYQYRAED---DTGCEAD-EITHLQATAQRYLHVYVGNRTASE 241  
DB 2026 IQVGSGDGESPAGRFTGVSTPDSGDTGDDAPREWTRHVS-----VLGEGDPATES 2078  
QY 242 TLPGLDGSAPSQADWLFYLVFDYBRSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFERT 301  
DB 2079 DHPGTDGD--GSAAM-----PPAAATATPLDGVYDLAELGYG--- 2115  
QY 302 RRLCRQVLMYHHLQALDLSKITEHNGPTLVSRLLINYDESAIASTLVFVR--RVGHEQDN 359  
DB 2116 -----GPAFOGLTGLWRDG---ADTLAEIRLPAQAQESAGL 2148  
QY 360 VVTLPP--LELAYQDFSPRHHAHQPMQVLANFNAIQR-----WQVLDKGEGLPGLLYQ 412  
DB 2149 FGVPALLDAALHPVLEGNSSAAGCD--AUTDATDRIRLPPAWAGVTLHAEGATAL--- 2203  
QY 413 DKGAWVRSRQRLBIGSDAVTWKMQ-----PLSVIPSLQSNASLVDINGD--GOLD--- 463  
DB 2204 -----RVRIITGPTVTLRLTDTTGAPVAIVESITLRAVAKDRLGTAGRVDDAL 2254  
QY 464 -----WVITG---PGLRGVHSQRPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSLVL 516  
DB 2255 FTVVWTGTTPPEAGRG-----AVEVE-----ELVDL--AGLGLDVE 2289  
QY 517 IGPKSVRILYAN--TRDGFAGKQDVQSGDITLVPFGADPRKLVAFSDVLGSGQ---AHVL 571  
DB 2290 LGAADVLRADRWTLT-----GDPSSAAARTAVRRTLAIVQEFSLSEPRFPGSLRV 2338  
QY 572 EVSARKVTCWNLGRGRFGQPIIT-----LPGFSQAPATEFNPAPQVYLADL--DGSQPTD-- 622  
DB 2339 CVTRGAVAALP-----GEDVTSIATGELWGLVRSQAENSENPGRLFLDLGEGEGERDGA 2391  
QY 623 -----LIYVHTNRLDIFLNKSGNGFAPVTLRP--PEGLRFDHTC 660  
DB 2392 BELIRATAGDEPQLAARDGRLLAPRLARTALSSEDTAGGAD-----RFGPDG-----T 2441  
QY 661 QLOMADVOGLGVASLILSVPHMSPHW--RCDLTWK-----PWLLENMNNNGVHTILRY 714  
DB 2442 VLVTGGTGGLGA-----LLARHLVHRVGRLLVSRRGADAPGAADLGEDLAGLGAEVAF 2497  
QY 715 RSSQFWLDEKAAALTT--GQTPCVLPFPPIHTLWQTE--TEDEISGNKLVTLR--YARG 769  
DB 2498 AARDAADRESLARATVPAEHP-----LTAVHTAGVDDATVEALTPERLDVLPKVD 2554  
QY 770 AWDGRE--REFRGFYBQTDHQL-----AQGN--APERTPPALTKWYATGLPVIDNAL 821  
DB 2555 AWWLHETKDLRLDAFLVLFSSVSGIVGTAGQANYAAANTGLDAAHRAATGLAA--TSL 2612  
QY 822 STEYWRDQAFAGP--SPRFTWQNKDVLTPEDDNSRYFNPNRALKQGLRSELYGLDD 879  
DB 2613 AMGLWDGTHGMGGTILGAADLARWSRAGITPLTP-----LOGLALFDAVAARD 2660  
QY 880 STNKHPYTVTFEFSQVRLQH--TDSRPVLWSSVSRNHYVIERIASDPQCSQNTLS 937  
DB 2661 AL--LVP-----AGLRTAHRGTGQPPALWRLVRRPRRAARTAE-----A 2702  
QY 938 SDRFQQLKQLSQVYP--RROQPAINLYPDTLPDKLLANSYDDQQLRLITYQQSSWHHLT 996  
DB 2703 ADTTGGWLSGLAAQSPSEERSTAVTLVTGVAD----- 2735  
QY 997 NNTVVLGLPDSSTRSDIPTYGAENVPAGLNLIELSKNSLIADKPREYLGQOKTAYTD 1056  
DB 2736 -----VLGHADSA-----AVGAER-----SFKD-----LGFDSLAGE 2763  
QY 1057 GQN-----TTPLOTPTQALIAETETTFNQSTLSAFN-----GSIPTDKLST--TLEQAGY 1106  
DB 2764 LFNRLNAATGLRLPA-----TTVFDHFPSPALASHLLAQVPLKEGTAATVVAE 2814  
QY 1107 QCTNYLFPRTGDKVVAHGVTDYGTAAQF-----WRPQKQNTQLTG--KITLI 1155

DB 2815 RGASFGDRATDDDDPIAIV-----GMACRYPGGVSSPEDLWRLVAETDRAISEPVRNG 2867  
QY 1156 WD-----ANYCVVVQTRDAAGLTTSYAKYDWRP--LTPVQLTIDINDNOHL--- 1197  
DB 2868 WDLBSLYDPDPESKGTTC-----REGGFLGAGDFDAAFGGISPREALVMDPQORLLE 2922  
QY 1198 ITLDALGR-----PITLRFWGTENGKMTGYSSPEKAS 1229  
DB 2923 VSWALERAGIDPSSLR--GSRGGVYVGAHGSYAS 2956

RESULT 8  
US-09-105-537-2  
; Sequence 2, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.  
; APPLICANT: Xue, Y.  
; APPLICANT: Zhao, L.  
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
; FILE REFERENCE: 600.438US1  
; CURRENT APPLICATION NUMBER: US/09/105.537A  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 5215  
; TYPE: PRT  
; ORGANISM: Streptomyces venezuelae  
US-09-105-537-2

Query Match 1.9%; Score 154; DB 2; Length 5215;  
Best Local Similarity 21.0%; Pred. No. 0.017;  
Matches 247; Conservative 123; Mismatches 426; Indels 380; Gaps 65;

QY 189 LEASVSSRGE---QIYYQYRAED---DTGCEAD-EITHLQATAQRYLHVYVGNRTASE 241  
DB 2026 IQVGSGDGESPAGRFTGVSTPDSGDTGDDAPREWTRHVS-----VLGEGDPATES 2078  
QY 242 TLPGLDGSAPSQADWLFYLVFDYBRSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFERT 301  
DB 2079 DHPGTDGD--GSAAM-----PPAAATATPLDGVYDLAELGYG--- 2115  
QY 302 RRLCRQVLMYHHLQALDLSKITEHNGPTLVSRLLINYDESAIASTLVFVR--RVGHEQDN 359  
DB 2116 -----GPAFOGLTGLWRDG---ADTLAEIRLPAQAQESAGL 2148  
QY 360 VVTLPP--LELAYQDFSPRHHAHQPMQVLANFNAIQR-----WQVLDKGEGLPGLLYQ 412  
DB 2149 FGVPALLDAALHPVLEGNSSAAGCD--AUTDATDRIRLPPAWAGVTLHAEGATAL--- 2203  
QY 413 DKGAWVRSRQRLBIGSDAVTWKMQ-----PLSVIPSLQSNASLVDINGD--GOLD--- 463  
DB 2204 -----RVRIITGPTVTLRLTDTTGAPVAIVESITLRAVAKDRLGTAGRVDDAL 2254  
QY 464 -----WVITG---PGLRGVHSQRPDGSTRFTPLNALPVEYTHPRAQLADLMGAGLSLVL 516  
DB 2255 FTVVWTGTTPPEAGRG-----AVEVE-----ELVDL--AGLGLDVE 2289  
QY 517 IGPKSVRILYAN--TRDGFAGKQDVQSGDITLVPFGADPRKLVAFSDVLGSGQ---AHVL 571  
DB 2290 LGAADVLRADRWTLT-----GDPSSAAARTAVRRTLAIVQEFSLSEPRFPGSLRV 2338  
QY 572 EVSARKVTCWNLGRGRFGQPIIT-----LPGFSQAPATEFNPAPQVYLADL--DGSQPTD-- 622  
DB 2339 CVTRGAVAALP-----GEDVTSIATGELWGLVRSQAENSENPGRLFLDLGEGEGERDGA 2391  
QY 623 -----LIYVHTNRLDIFLNKSGNGFAPVTLRP--PEGLRFDHTC 660  
DB 2392 BELIRATAGDEPQLAARDGRLLAPRLARTALSSEDTAGGAD-----RFGPDG-----T 2441

```
Qy 661 QLOMADVQGLGVASLILSVPHMSPHHW--RCDLTNMK-----PWLINEMNNMGVHHTLYR 714
Db 2442 VLVGTGGLGA-----LLARHLVERHGVRLLLSRRGADAPGAADLGEDLAGLGAEVAF 2497
Qy 715 RSSQFWLDEKAAALTT--GQTPVCYLFPFPIHTLWQTE--TEDEISGNKLVTTILR-YARG 769
Db 2498 AAADAADRESLARAIAATVPAEHP--LTAVVHTAGVVDATVEALTPERLDAVLRPKVDA 2554
Qy 770 AWDGRE--REFRGYVQETDISHQL-----AQGN--APERTPPALTKWYATGLPIDNAL 821
Db 2555 ANWLHETKJLRLDAFLVFSVSGIVGTAGQANAAANTGLDAAHRAATGLAA--TSL 2612
Qy 822 STEYWRDQAPAGF--SPRFTTWQNDKDVLPITPDDNSRYWFNKALGQLRLSRLYLGLDD 879
Db 2613 AWGLWDGTHGCGTILGAADLARWSPAGITPLTP-----LQGLALFDAVAARD 2660
Qy 880 STNKHVPYVTFERSQVRLQH--TDSRYPVLMSVSVESRNYHYERIASDPQCQNITLS 937
Db 2661 AL--LVP-----AGLRPTAHRGTGQPPALWRGLVRAARPRRAARTAAE-----A 2702
Qy 938 SDRFGQPLKQLSVQYP-RRQOPAINLYPDTLPDKLLANSYDDQQRLRLTYQQSSWHHLT 996
Db 2703 ADITGGWLSGLAQSPEERSTAVTLVTGVVAD----- 2735
Qy 997 NNTVRVLGLPDSRSDIFTYGAEVNPAGLNLLELLSKNSLIADDKPREYLGQOKTAYTD 1056
Db 2736 -----VLGHADA-----AVGAER-----SFKD-----LGFDSLAGE 2763
Qy 1057 GQN-----TTPLOTPTQALIAFTFTVNQSTLSAFN-----GSIPSDKLST--TLEQAGY 1106
Db 2764 LRNLNAAATGLRLPA-----TTVPDHPSPAALASHLLAQVPGLEKGTAAATVVAE 2814
Qy 1107 QQTNYLPRTEGDKVVAHGYDYGTAAQF-----WRPQKQNTQLTG-KITLI 1155
Db 2815 RGASFGDRATDDPIAIV-----GMACRYPGGVSSPEDLWRLVAEGTDAISBFPVNRG 2867
Qy 1156 WD-----ANYCVVVQTRDAAGLTTSAKYDWRP--LAPVQLTDINDQHL--- 1197
Db 2868 WDLESLYDPDESQKTTYC-----REGGFLEGAGDFDAFTGISPREALVMDPQORLLE 2922
Qy 1198 ITLDALGR-----PITLRFWGTENGWMTGYSSPEKAS 1229
Db 2923 VSWEALERAGIDPSSLR--GSRGGVYVGAAGHSVAS 2956

RESULT 9
US-09-543-681A-7181
; Sequence 7181, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7181
; LENGTH: 1028
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-543-681A-7181

Query Match 1.9%; Score 153.5; DB 2; Length 1028;
Best Local Similarity 18.0%; Pred. No. 0.001;
Matches 204; Conservative 124; Mismatches 356; Indels 449; Gaps 52;

Qy 19 GAITGMGALFTGPDGMAAL-----SLPLPTISAGRGYAPATFLNYSAGHSPPGLG 71
Db 162 GSATGKLDKLVITAPDQAGTLFFDQIIMSVELD---NRWPIEDYQIPYVNNVNTVMSKN 218
```

```
Qy 72 WDCNMVTIRRRTHRGVPHYDETDTFLGPEGEVLVADQPRDEBTLQINLGATFTVTGYR 131
Db 219 WSALLM-----YDQMFTAHP--TLNFATEFRDD-----Q 246
Qy 132 SRLSHFSRLBYWQPKTKTDFWLIYSPGQVHLLGKSPQARISNFSQTTQTAQWILLEA 191
Db 247 PEVASIYQRFPEYQ-----GVSRRDKKITAEMIDKNLALM--- 280
Qy 192 SVSSRSGQIYYVYRAEDDTGCEADEITHHLQATQRYLHVIVVYGNRTASETLPGLDGSAP 251
Db 281 -----KXGALAEQHA-----DSGIT 294
Qy 252 SQADWLFYLVDFDGERSNNLTKPPAFSTGTGSLWCRODRFSRYEYGFIRTRRLCRQVLMY 311
Db 295 AKA-----LDHNRQNRFKVEGVFSE-----ETQK----- 319
Qy 312 HHLOALDSKITEHNGPTLV-----SRLILNVDESAIASTLV-----FVRVUGH 354
Db 320 ---ALLDANMLRDVGKTLLOTAIYLRSRSLASIDRKKLETLYLLGTRYVLEQGFTRGSGY 376
Qy 355 EQDGNVVTLPLELAYOD-----FSPRHHAHQPMQMDVLANFNAIQRWQLVDLKGELP 407
Db 377 Q-----IIT-----HVGYOTRELFDAWFTGRH-----ILAKHMLLAPTQ----- 410
Qy 408 GLLYQDKGAWYRQAQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVYDINGD---GOLDW 464
Db 411 -----QAMWYNATGRI-----FEKDNEIVDANVDILNTQLQW 443
Qy 465 VITQFGLRGVHSQRPDG-----SWTRFTPLNALPV-----EYTHPRAQLADLM 507
Db 444 MIKSLMLPDYQQRQALAOQLQSLWLNKTIILSSKGAVGGFKSDGSIFFHSHYPAYAKDAF 503
Qy 508 GAGLSDLVLIGPKSVRL---YANTRDGFAGKDVQVQSGDITTLPPVG-----ADP 553
Db 504 GGLAPSYALSHSPFLRSSPAHARKOVLLKMRITYKTETQIPLVLSRHPGLHKISIDP 563
Qy 554 RKLVAFSDVLGSGQAHILVEVSATKVTCPNLR-----GRFGQPIITLPQFS 599
Db 564 FKWMALAGTPDGKQ---ELDTTLAAAYAKLANKDSPEGIKAENEYPVGAWAMNYASMAIQ 619
Qy 600 QPATEFNPAQVYLADLDGSGPTDLI-----YVHTNRLDIF---LNKSG----- 639
Db 620 RRASITAPQOSWLAIAARGFS-RYLIVGNESYENNRRYGRYLOYGQLEIIPADLTKSGFSHA 678
Qy 640 ---NGFAEPVTLRFPEGLRFDHTCLOQADVOGLGVASLILSVPHMSPHHWCDLTNMK 695
Db 679 GWDWNRYPGTTTTIHLF---YDE-LEAKLSQLPSAGIEEMLLSTQRYSG----- 722
Qy 696 PWLLNEMNNN---MGVHHTLYR-----SSSQFWLDEKAAALTTG-----QTFVC 737
Db 723 ---ANTLNNNSMFAMKHLGHSKYQQSLRANKSYFLFDNRVIALGSGIENNDKQHTTETT 779
Qy 738 YLPPPIHTLWQTEDEISGNKLVTTILRYARGAWDGRERFRGRGYVEQTDHSHOLAQNA 797
Db 780 LFQFAVPKL-----QSIINGKKVQNQ-----GTQULTLNA 810
Qy 798 PERTPPALTKWYATGLPIDNALSTYWRDD-----QAFAP-----GFSPEFTTWQD 844
Db 811 DTLIDPAGNLYKLAKG-QTFEFSYQKQYVDDNRNSQQOTEQLFATAVISHGKAPK----- 863
Qy 845 NKDVPLTPEDDNRSYWFNKALGQLRLSRLYGLDDSTNKHVPYVTFEBSQVRLQHTDS 904
Db 864 -----NANYEYAIAEAQ-----DNKAPEYTV-----LQHNQ 891
Qy 905 RYPVLMSVSVESRNYHYERIASDPQCQNITLSDRFGQPLKQLSVQYPRQO-----QP 958
Db 892 LHAVK-DKITQEGYAFPN-ATEVNSSQALLSSD---SP-TMWVXKQKQKQLTSLVNP 945
Qy 959 AINLYPDTLPDKLLANSYDDQORLRLTYQQSSWHHLT-----NNTVRVLGL 1005
Db 946 DLNLYQGIADQI-----DNKGNQVEVSYSRQW--LTADPQPISSIVTVKGI 991
```

RESULT 10  
 US-09-711-164-467  
 ; Sequence 467, Application US/09711164  
 ; Patent No. 6589738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Forsyth, R. Allyn  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY  
 ; FILE REFERENCE: ELITRA.008A  
 ; CURRENT APPLICATION NUMBER: US/09/711,164  
 ; CURRENT FILING DATE: 2000-11-09  
 ; PRIOR APPLICATION NUMBER: US 60/164415  
 ; PRIOR FILING DATE: 1999-11-9  
 ; NUMBER OF SEQ ID NOS: 469  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 467  
 ; LENGTH: 1377  
 ; TYPE: PRT  
 ; ORGANISM: Escherichia coli  
 US-09-711-164-467

Query Match 1.8%; Score 144; DB 2; Length 1377;  
 Best Local Similarity 18.6%; Pred. No. 0.012; Mismatches 403; Indels 472; Gaps 62;  
 Matches 230; Conservative 133;

QY 406 LPGLLYQDKGAWYRSAGRLGIGSDAVTWKMQPLSVIPSLQSNASLVVDINGDGLDWV 465  
 DB 70 LPPILSRYSYRTKTPAPVGLSGP--WK--MPADIRLQRLDNTLLISDNGGRSLYFE 124  
 QY 466 ITGPGLRGHSRPPGSGWT-----RFTPL-NALPVEY-THPRAQLAD----- 505  
 DB 125 HLPFGEDGY--SRSESLVLRGVAKLDEGHRLAALWQALPEELRLSPHYLATNSPOQP 182  
 QY 506 --LMG-----AGLSDLVILGP-KSVRLYANTRDGFAGKDVQ-----SGDITLVPQAD 552  
 DB 183 WLLGWCERVPBEADVLPAFLPPYVLTGLVDRFGRTQTFHREAAAGEFSGEITVTDGAW 242  
 QY 553 PR-KLVAFSDVLGSGOAHLEVS-ATKVTCPNLCGRFGQPIITLPGFSQAPATEFNPQV 610  
 DB 243 RHFRVLVITTOQARAEARQALSGGTESAFPD-----TLPGYTE----- 282  
 QY 611 LYADLDGSGPTDLIYVHTN--RLD-IFLNKSGNGFABPVTLFPPEGL-----RFDHTCQ 661  
 DB 283 -----YGRDNGIRLSAVWLTH-----DP---EYPENLPAAPLVRYGWTPT 319  
 QY 662 LQMA---DVQGLGVASLILSVPH---MSPH-----HWRCDLTMKFWLLNEMNNMG 707  
 DB 320 GELAVVYDRSGQVRSFTYDDKYGRMVAHRHTGRPEIRYRYSDBR-----VTEQLNPAG 375  
 QY 708 VHTLRY---RSSQFWLDEKAAALTGTQTPVCYLPPIHTLWQTEDEISGNKLVTL 764  
 DB 376 LSYTYQEKDRITITDLSRREVLTQGE-----AGLRVVK 413  
 QY 765 RYAGAWDGREREPFGYVE-OTDSHQAQGNAPERTPPALTKNWYATGLPVIDNALST 823  
 DB 414 EHADGS--VTQSQFQDVGRLAQTQD---AAGRTTEYSDDVT--GLITRIITPDGRASA 465  
 QY 824 EYWRDDQAPAGSPRETTWQDNKV-----PLTPEDDNRVWFNPAKGLQLLR 871  
 DB 466 FYNNHNLQTS-----ATGPDGLELRREYDEGLRIQETAPDGDITRY----- 510  
 QY 872 SELYGLDSTNKHVPVTVTFRSQVRRLQHTDSRPVLMWSVVESSRNTHYERIASDPOCS 931  
 DB 511 -----DNPHSDLP-CATE-----DATOSRKTMTWS-----RYGQLLSFTDCS 546  
 QY 932 QNIT-LSSDRFGQ-----PLKQLSVQYPRQOPAINLYPDTLPDLKLLANSYDDQORQ 982  
 DB 547 GVTVTRYDHRFGQMTAVHREGLSQVRAVDSRGLIAVK-----DTQGE 591  
 QY 983 LRLTYQSSWHHLTNNTVRLGLPDSTRSDIFTYGAENVPAAGLNLELLSDKNLSIADK 1042  
 DB 592 TRYEY-----NIAGDLTAVIAPDG 610

QY 1043 PREYLGQOKTAY-----TDQNTTPTLQTPTRQALIAFTTTFVFNQSTLSAFNGSIPS-- 1094  
 DB 611 SRN--GTQYDAWGKAVRTTQGLTRSMEDVDAAGRVIRLTSE-----NGSHTTFR 657  
 QY 1095 -DKLSTTLEQAGY--QOTNYLFPRTG-----EDKVVVAHHGY-----TDYGTAAQF 1137  
 DB 658 YVDLRLIQTGFGDRTQRYHDLTGKLIIRSEDEGLVTHWHYDEADRLTHRTVKGETAER 717  
 QY 1138 WRPOKQ-----SNTQLTGK-----ITLIWD----- 1157  
 DB 718 WQYDERGLWLTISHISEGHRVAVHYRDEKGLTGERQTVVHPQTEALLWQHETRAYNA 777  
 QY 1158 ---ANYCVVQTRDAAGLTTSKAYDMRFLTPVQLTD-----INDNQHLITIDALGR-PI 1207  
 DB 778 QGLANRCIPDSLPVAVELT---YSGYLAGMKLGDTPLVEYTRDLRLHRETLRSFGRYEL 833  
 QY 1208 TLRP-----WGTENGKWTGYSSEKA---SFSPSPDVNAALIELK 1243  
 DB 834 TTAYTPAGQLOSHLNSLLSDRYTM-NDNGELIRISSPRQTRSYSTTGLTGVHHTA 892  
 QY 1244 KPLPVAQCVVAPESWMPVLSOKTFNRLAEQDWQKLYNARIITEDGRICTLAYRRWVQSQ 1303  
 DB 893 ANLDIR-----IPYATDPAGNRLPD----- 912  
 QY 1304 KAIPOLISLANNPRLPPHS-LTLTLD-RYDHPDQIQVQVFSDFGRLLQAAARHEA 1361  
 DB 913 -----PELHPDSTLSMMPDNRIARDAHYLYR-----DRHGLRTEKTDLIPE 954  
 QY 1362 GMARQBNEDGSLIINVQHTENRWATGTEVDNKGQPIRTQPYFELNDWRYVNSDSAROE 1421  
 DB 955 GVIRTDDE-----RTHRYHYDSQHRLVHYHTRTQYE 984  
 QY 1422 KEAVADTHVYDPIGRIKIVITAKGWFRTLTFTPMFTVN 1459  
 DB 985 EPLVESRILYDPLGR---VAKRWRRERDLTGWMSLS 1019

Query Match 1.8%; Score 143; DB 2; Length 798;  
 Best Local Similarity 20.9%; Pred. No. 0.0056;  
 Matches 137; Conservative 81; Mismatches 232; Indels 206; Gaps 34;  
 QY 18 GGAITGMEALTPGPDGMAALSLPLPISAGRGYAPAFNLNYSAGNSPFLGWDC--- 74  
 DB 123 GYNIRGMESNRVGDVQIAQ-----PNATGRGV-----GRAGLNTFGIGRDYIDP 169  
 QY 75 -----NMTIRRTHEGV-PHYDET----- 94  
 DB 170 YWYGSVDIQSGATSTETANSAGNVSPRPSADYLPFGKTSFAGSYGSDASRSMHN 229  
 QY 95 --TFGLPGEVL--VVADQPRDESLQGINLGATFTVGYRSLRLESH-FSRLEYWQP--- 146  
 DB 230 GVTVAGGD-BFLRGILVYRRDRDQETEN-NSG---TVDAYPANMHSDAFLASGIWQPNDE 284

RESULT 11  
 US-09-489-039A-10045  
 ; Sequence 10045, Application US/09489039A  
 ; Patent No. 6610836  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gary Breton et. al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 2709.2004001  
 ; CURRENT APPLICATION NUMBER: US/09/489,039A  
 ; CURRENT FILING DATE: 2000-01-27  
 ; PRIOR APPLICATION NUMBER: US 60/117,747  
 ; PRIOR FILING DATE: 1999-01-29  
 ; NUMBER OF SEQ ID NOS: 14342  
 ; SEQ ID NO 10045  
 ; LENGTH: 798  
 ; TYPE: PRT  
 ; ORGANISM: Klebsiella pneumoniae  
 US-09-489-039A-10045



Db 731 --DDKNDQGRPEKVS-----VNLADGE---KVKTLDVTS-ETNWKYEFKDLPKYD 777  
Qy 1184 --TPVQLTIDINDNOHLITLIDALGRPITLIRFWGTENGKMTGYSSEPKASFPSPDVNAAIE 1241  
Db 778 EGKLEYVTEDHVXDYTTDINGTITNKY-----TPGETSAIV- 816  
Qy 1242 LKKPLPVAQCQVYAPESWMPVLQKTYFNRLAQDWOKLYNARIITEDGRICTLAYRRWQ 1301  
Db 817 -----TKNWDNNN-----QDGKRPTEIKVELYQ 840  
Qy 1302 SOKAIPQLISLANNPRLPHSLTLTDRYDHPQOIRQOVVSDGFGRLLOAAARHEA 1361  
Db 841 DKGATGK-TAILNES-----NNWTHWTGLD---EKAGQQVKYTVBELTKVGYTHV- 890  
Qy 1362 GMAQRNEDGSLIINVQHTENRWAVTGRTEYDNK-----GQPIRTYQ 1403  
Db 891 -----DNNDMGNLIVTNKYTPETTSISGEKVDDKNDQGRPEKVSNNLANGEKVKTL 946  
Qy 1404 PYFLNDWRYVNSDSARQEK-----EAYADTHYDPIGREKIVITAKGWFRTLTPT 1454  
Db 947 VTSETNWKYFADLPKYDEGKIEYVTEDHVKDVTYTDINGTITNKYTPGETSATVTKN 1006  
Qy 1455 WFTVNEDENDTAAEVK 1470  
Db 1007 WDDNNQDGRKRPTEIK 1022

RESULT 13

US-08-346-455B-38  
; Sequence 38, Application US/08346455B  
; Patent No. 5731167  
; GENERAL INFORMATION:  
; APPLICANT: UNITED STATES OF AMERICA; DEPT.  
; APPLICANT: OF HEALTH AND HUMAN SERVICES  
; TITLE OF INVENTION: MOTILITY STIMULATING  
; TITLE OF INVENTION: PROTEIN USEFUL IN CANCER DIAGNOSIS AND  
; NUMBER OF SEQUENCES: 69  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORGAN & PINNEGAN  
; STREET: 345 PARK AVENUE  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/346,455B  
; FILING DATE: 28-NOV-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06613  
; FILING DATE: 24-MAY-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/249,182  
; FILING DATE: 25-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/822,043  
; FILING DATE: 17-JAN-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DOROTHY R. AUTH  
; REGISTRATION NUMBER: 36,434  
; REFERENCE/DOCKET NUMBER: 2026-4149PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 38:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 979  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: Unknown  
MOLECULE TYPE: protein  
HYPOTHETICAL: No  
ORIGINAL SOURCE:  
ORGANISM: Human  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE: Liver  
CELL TYPE:  
CELL LINE:  
ORGANELLE:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION: putative autotoxin  
OTHER INFORMATION: protein sequence from human liver  
US-08-346-455B-38  
Query Match 1.8%; Score 140; DB 1; Length 979;  
Best Local Similarity 18.9%; Pred. No. 0.015;  
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;  
Qy 418 WYSAORLGEIGSDAVTWKMQPLSVIPSLQSNASLVINGDGLQDWI-----TGPGLRG 473  
Db 261 WITATKORGE-----SWN-----ILLVCCPSRAELIT-----LQWTLDPHERPSVYA 305  
Qy 474 YHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMA-----GLSDLVLIIGPKSVRLYANTRD 530  
Db 306 FYSEQPDFSGHKMFEGP---EMTNPLREMKHIVQQLMDGLKQLKLRVCNV-IFVETMD 361  
Qy 531 GFAK-----GKDVQSGDITLVPGA-----DPRKLVAFSDVLGSGQAH 569  
Db 362 GRCHMYRTFLSNLYLTNVDDITL-VPGTLGRIRSKFSNNAKYDPKAIIA----- 409  
Qy 570 LVEVSATKVTCPNPLGRGRFGQPTLPGFSOPATEFNEPAQVVLADLDGSGPTDLIYVHTN 629  
Db 410 -----NLTC-----KKPDQHFKE---YLKQ---HLPKRLHYANNR 438  
Qy 630 R-----LDIFLNKSGNGFAEPVTLRPFEGELRPHDHTCOLQMAADVQGLG 671  
Db 439 RIEDIHLLVERRHVARKPLDVKYKPSGNAFSRETT-----AFDNKVNMQTVFVGVG 491  
Qy 672 -----VASLILSVPHMSPHWECDLTNMKPWLNNMNNMGVHHTLRYRSSSQFWLDEKAA 727  
Db 492 PTFKYTKVPPFENIELYNVNCDDLGLKP-----APNNGTHGSLNHLRLTNTFRPTMPE 545  
Qy 728 ALTTGQTP-VCYLPFP-----IHTLMOTETEDSISGNKLVTT 763  
Db 546 EVTRPNYPGIMYLOSDFDLGCTCDDKVEPKNKLDELNKLRLTKGSTERHLLYGBRPAVL 605  
Qy 764 LR-----YARGAWDGRERERFGEYVEQTDSDHQLAOGNAPE-----RTPPALTK 807  
Db 606 YRTRYDILYHTDFESGYSEIEFLMPLWTSYTSYKQAEVSSVPDHLTSCVRPDRVRSFSQ 665  
Qy 808 NYATGLPVIDNALSTEWRRDDQAFAG-----SP--RFTTWQDNKDVPLTDEDDNS 857  
Db 666 NCLA-----YKNDKQMSYGLFPFPPYLSSSPEAKYDAPLVNTMVPMPYAFKRV 712  
Qy 858 RYWFNRAL-----KQOLLRESELYGLDDSTNKHVPYTVTFERSQVRLQHTD 903  
Db 713 WNYFORVLVKKYASERNGVNVISGPIFYDYDGLHDTEDKIKQYV-----EG 759  
Qy 904 SRYPVLMSSVVESRNYHYERIASPPQCSQNTLSSDRFGQPLKQLSVQVPRRQOPAINLY 963  
Db 760 SSIPV-----PTHYYSIITS-----CLDTPQADKCDGPLSVSSFILPHR----- 799  
Qy 964 PDTLPDKLLANSYDDQQRQLRLTYQQSSW---HLLTNTNTRVVLGLPDSRSDIFTYGAENV 1021





FILE REFERENCE: 2026-4149US4  
CURRENT APPLICATION NUMBER: US/09/483,831B  
CURRENT FILING DATE: 2000-01-17  
PRIORITY APPLICATION NUMBER: 07/822,043  
PRIORITY FILING DATE: 1992-01-17  
PRIORITY APPLICATION NUMBER: 08/249,182  
PRIORITY FILING DATE: 1994-05-25  
PRIORITY APPLICATION NUMBER: 08/346,455  
PRIORITY FILING DATE: 1994-11-28  
PRIORITY APPLICATION NUMBER: 08/977,221  
PRIORITY FILING DATE: 1997-11-24  
NUMBER OF SEQ ID NOS: 70  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 70  
LENGTH: 979  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Putative autotoxin protein sequence from human  
OTHER INFORMATION: liver.  
NAME/KEY: VARIANT  
LOCATION: (860)  
OTHER INFORMATION: Xaa at positions: 860, 889, 905, 911, 927, 937,  
OTHER INFORMATION: 944, 950, 954, 967, and 975 represents an unknown  
OTHER INFORMATION: or other amino acid.  
US-09-483-831B-70

Query Match 1.8%; Score 140; DB 2; Length 979;  
Best Local Similarity 18.9%; Pred. No. 0.015;  
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;

QY	418	WYSAQRLEIGSDAVTWKMQPLSVIPFSLQSNASLVINDGQGLDWI----	TPGGLRG	473
DB	261	WITATKQGE-----SWN-----ILLVCHPSRAEILTI-----LQMLTLPDHERPSVYA	305	
QY	474	YHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMCA-----GLSDLVLIGPKSVRLYANTRD	530	
DB	306	FYSEQPDFSGHKMPFGP---EMTNPLREMHKIVGQLMDGLKQLKLRGVNV-IFVETMD	361	
QY	531	GFAP-----GKDVVQSGDITLPVGA-----DPRKLVAFSVDVLGSGQAH	569	
DB	362	GRCHMYRTEFLSNLTNVDDITL-VPGTLGRIRKFSNNAKYDPAIIA-----	409	
QY	570	LVEVSATKVTCPNIGRGRFGQPIITLPGFSQPATEFNPQAQVYLADLDGSGPTDLIYVHTN	629	
DB	410	-----NLTC-----KKPDQHKP---YLKQ---HLPKRLHYANNR	438	
QY	630	R-----LDIFLNKSGNGFAEPVTLRFPPEGLRFDHTCQLQOMADVQGLG	671	
DB	439	RIEDIHLLVERRHWVARKPLDVYKPKSGNAPSRETT-----AFDNKVNMQTVFVGYG	491	
QY	672	---VASILSVPHMSPHHWRCDLTNMKPILLNENNNMGVHHTLRYRSSQFWLDEKAA	727	
DB	492	PTFKYTKVPPENIELYNVCMDDLGLRP-----APNNGTHGSLNHLRLRTNFTPTWPE	545	
QY	728	ALTTGQTP-VCYLPFP-----IHTLMOTETEDIEISGNKLVTT	763	
DB	546	EVTRNYFGIMYQLQSDFDLGCCTDDKVEPKNKLDELNKLHTKSGSTEERHLLYGRPAVL	605	
QY	764	LR-----YARGAWDGRREFRFGYVEQTDSHQLAQNABE-----RTPPALTK	807	
DB	606	YRTRVDILYHTDFESGYSEIFLPLMTSVTSKQAEVSVDPDLTSCVRPDPVRVSPSPSQ	665	
QY	808	NWYATGLPVIDNALSTYWRDDQAFG-----SP--RFTTWQDNKDVPPLTPEDDNS	857	
DB	666	NCLA-----YNDKQMSYGFLLPFPYLSLSSPEAKYDAFLVTNMVPMYPAFKRV	712	
QY	858	RYWFNRAL-----KGQLLRSELGLDDSTNKXVPYTVTFEFSQVRRLOHTD	903	
DB	713	WNYFORVLVKYASERNGVNIVSGIFDYDYDGLDHTEDKIQYV-----EG	759	
QY	904	SRYPYLWSSVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQLSQVQPRRQPAINLY	963	

Db 760 SSIPV-----PTHYYSIITS-----CLDFTQPADKCDGFLSVSSFILPHR----- 799  
QY 964 PDTLPDKLLANSYDDQQRQLRLTYQSSW--HHLTNNTVRVLGLPDSTRSDIFTYGAENV 1021  
Db 800 FD--NEESCNSSED-----ESKWVEELMOMHTARVRDIEHLTSLDFFRKTSSY 846  
QY 1022 PAGGLNLELLSDKNSL 1037  
Db 847 P-----EILTKTYL 856

Search completed: February 16, 2006, 21:46:34  
Job time : 40.2035 secs

**THIS PAGE BLANK (USPTO)**

---